UNDERSTANDING THE CO-EMERGENCE OF URBAN LOCATION CHOICE AND MOBILITY PATTERNS

EMPIRICAL STUDIES AND AN INTEGRATED GEOSPATIAL AND AGENT-BASED MODEL

Ransford Antwi Acheampong
Jesus College

Department of Land Economy
University of Cambridge

This dissertation is submitted
for the degree of Doctor of Philosophy

April 2017

AUTHOR DECLARATIONS

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text.

During the period of registered study in which this thesis was prepared, the author has not been registered for any other academic award or qualification. The material included in this thesis has not been submitted wholly or in part for any award or qualification other than that for which it is now submitted.

This dissertation does not exceed the regulation length, including footnotes, references, and appendices.

Ransford Antwi Acheampong
April 2017

ACKNOWLEDGEMENTS

I owe much gratitude to my supervisor, Dr. Elisabete A. Silva for introducing me to the field of Urban Modelling and for spending time to review my work; I am grateful for all the feedback that has gone into making this research a success.

I have benefited immensely from the opportunity to interact and share ideas with faculty members and colleague researchers on various platforms during this journey. In particular, I wish to thank all members of the Agent-based Modelling Group and the Geo-Information Science Forum, based in the Department of Geography, Cambridge for providing a focal point to gain very useful methodological insights for my research.

This PhD experience would not have been possible without financial support. I wish to express sincere gratitude to the Cambridge Commonwealth, European and International Trust for graciously providing substantial funding under the Cambridge International Scholarship Scheme (CISS) to support successful completion of this research.

I had a great time at Jesus College. I would like to thank the college staff for providing a supportive environment for all the period of residence there. All financial grants received from the college in support of my studies is gratefully appreciated.

Finally, I would like to thank my office colleagues, Rosemary Ostfeld, Pawel Kransy and Goyllete Chami for their support and encouragement, and fellow members of the Lab of Interdisciplinary Spatial Analysis (LISA) at the Department of Land Economy, Felix Agyemang, Lun Liu, Chaowei Xiao, Bingyu Zhao and Gerrard Cassey for their support in diverse ways.

To

My wife, Nana Adwoa Nyarko

What will I do without your unconditional support and encouragement?

Intentionally Left blank

ABSTRACT

Understanding and simulating the relationship between urban land-use configuration and patterns of human spatial interaction has been the subject of multi-disciplinary research. Conceptually, it is recognized that the location decisions of several urban actors including individuals, households, firms and public sector institutions, collectively determine the spatial distribution of land-use activities; the emergent land-use patterns, in turn, provide the structural conditions within which flows and interactions between locations occur daily and respond to each other over time. Over the past six decades, various theories and concepts from urban economics, social-physics, transportation studies, and the complexity sciences have underpinned empirical research and development of state-of-the-art simulation models to explore the land-use and travel nexus.

Using a case study design and selecting the Kumasi Metropolis, a medium-size metropolis of nearly two-million inhabitants in Ghana, West Africa as the case study area, two main objectives, which reflect research trends and gaps in both the empirical literature and simulation model development have been addressed in this thesis. The first objective was to examine empirically, the location choice behaviour of households and individuals with respect to their residential and job locations, and the mobility patterns associated with the observed home-work location combinations within the metropolis. The second objective was to develop an integrated geospatial and agent-based model to simulate how the residential and job location choice behaviour of heterogeneous households and individuals co-emerge with mobility patterns in the metropolis.

The empirical studies presented in this thesis contributes to a deeper understanding of how location-defining attributes at multiple spatial-scales interact with socio-demographic attributes of heterogeneous households and individuals to determine their residential location choice, job location choice and mobility characteristics. The development of the Metropolitan Location and Mobility Patterns Simulator (METLOMP-SIM)—an integrated geospatial and agent-based model also demonstrates how the encoded micro-scale behaviour of purposive households and individuals, interacting with each other and their environment dynamically, could reproduce macro-scale urban location patterns, property market price formation and evolution, and patterns and attributes of spatial flows and interactions anchored on the population's residential-job location combinations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: GENERAL INTRODUCTION

1.1 Background to research and motivation

The relationship between urban land use configuration and patterns of human spatial interaction has been the subject of research since the pioneering work of Hansen (1959) in Washington DC. Hansen's original work demonstrated that urban activity locations and mobility patterns co-determine each other. At a conceptual level, the decisions of several urban actors including individuals, households, firms and public sector institutions over time, collectively determine the spatial distribution of land use activities; the emergent land-use patterns reflect not only the location of urban activities such as the place of residence, employment and supporting facilities but also the densities, diversity and accessibility of activities, which provide the structural conditions within which flows and interactions between locations occur daily (Wegener and Fuerst, 2004; Næss, 2015; Aditjandra et. al, 2013).

Over the past six decades, various theories and concepts from urban economics, urban geography, social-physics, transportation studies, behavioural economics and the complexity sciences have underpinned research on the urban structure and mobility patterns nexus across different disciplines. Broadly speaking, research in the field has focused on two mutually dependent areas. The first dimension of research, taking empirical focus, have sought to explore and understand the processes underlying the emergence of urban structure and how urban structural characteristics impact patterns of spatial interaction and vice versa. The second aspect of research efforts have gone into the development of state-of-the-art urban land use and transport simulation models for both exploratory research and practical policy application purposes.

While the influence of urban structural variables such as density, land-use diversity and destination accessibility on travel behaviour has received considerable attention in the empirical literature (e.g. Gim, 2013; Eboli et. al, 2012; Grunfelder and Nielsen 2012; Ewing and Cervero, 2010), studies of the long-term urban location choice behaviour of individuals and households with respect to residence and employment remain fundamental to understanding the relationship between urban land use and mobility patterns. This is because from a bottom-up perspective, the residential location behaviour of heterogeneous households and the job location choice of individuals, interacting with the location decisions of firms and public sector institutions' formal policy responses in the domains of urban land use planning, underline the emergent structure of cities (Batty, 2013; Pagliara et al., 2010; Lee and Waddell, 2010). The ability to understand and model the urban location choice process is therefore fundamental to understanding the long-term choices that condition daily mobility patterns, improve long-term travel demand forecasting, and inform the design of built environments that can shape desirable travel behaviours (Habib and Kockelman, 2008; Pinjari et al, 2011).

In the domain of simulation model development, state-of-the-art prototype and operational models, drawing on the empirical insights, the various theoretical proportions and simulation paradigms, have been developed to explore the processes underlying the co-evolution of urban spatial structure and travel demand and for practical policy decision-making purposes. As will be elaborated later in this thesis, three main types of models according to purpose can be identified in the literature. These are urban land-use models developed to simulate location choice (e.g. Zhuge et al. 2016; Tannier et al., 2016; Ettema, 2011; Filatova et al, 2011); urban models that focus exclusively on modelling travel demand (e.g. Eluru et al. 2008; Arentze and Timmermans 2004) and operational land use and transport interaction (LUTI) models such as UrbanSim (Waddell et al., 2003), MARS (Pfaffenbichler, 2011) and TRESIS (Hensher and Ton, 2002) which integrate dynamically, the feedback relationship between the land use and transportation systems.

Notwithstanding the considerable amount of research, the literature points to a number of interesting fundamental empirical questions as well as simulation model development opportunities worthy of pursuit to extend the current body of knowledge in the field.

Firstly, the need for the accretion of more fundamental empirical insights about the location choice behaviour of households and individuals in different contexts is emphasized in the literature. Of importance is the need for research to move beyond the traditional approach where residential location choice studies have presented aggregate spatial zones as discrete choice alternatives (e.g. Pinjari et al., 2011; de Palma et al., 2007; Bhat and Guo, 2007; Pagliara et al.,, 2010) towards new approaches where attributes defining location at the meso and macro scales as well as attributes of dwelling units at the micro-scale (e.g. Habib et al., 2011; Lee and Waddell, 2010; Bhat, 2015) are represented. In addition, the need to verify empirically, the interdependence between residential and job location choice among individuals and within multiple worker households so as to offer robust understanding beyond the exogenous work-place principle grounded in the traditional assumption of monocentric urban structure, is emphasized in the literature (Lee and Waddell 2010; Yang et al., 2013; Habib et al., 2011).

The second opportunity for research advancement in the field of urban location choice modelling is methodological. Traditionally, classical urban micro-economic theories (e.g. Alonso, 1964; Wingo, 1961) and entropy-based spatial interaction approaches grounded in the Newtonian concept of gravity (Lowry,1964; Wilson, 1970) have provided the standard reference point for understanding and modelling urban location choice. In recent years, the need for theory-driven disaggregate modelling approaches capable of capturing the complexities of location choice behaviour at the individual level while overcoming the weak assumptions of their predecessors has been emphasized in the literature (Acheampong and Silva, 2015; Rasouli and Timmermans 2014a).

Agent-based modelling (ABM) constitutes one of the innovative methodological advances towards the development of fully disaggregate, stochastic models. Adopting a bottom-up computational paradigm, the approach allows for the creation, analysis and experimentation with models composed of autonomous agents that interact with each other and their environment locally (Railsback and Grimm, 2011; Wu and Silva, 2010; Batty, 2007).

In the last decade, a new generation of disaggregate models of urban location choice employing ABM paradigm have been developed (e.g. Ettema, 2011; Lemoy et al., 2010; Filatova et al., 2011; Magliocca et al. 2011). However, these models differ considerably in their ability to realistically capture the essential aspects of the phenomenon they simulate and in their outputs capabilities. As will be elaborated later in chapter two of this thesis, the clear majority of these new generation of disaggregate models do not incorporate the full range of real-world market dynamics including bilateral transactions and competitive bidding that shape price formation as well the range of discrete choice alternatives and location-defining attributes at the different spatial resolutions (e.g. Tannier et al., 2015; Hosseinali et al., 2013; Murray-Rust et al., 2013; Brown et al., 2008).

Moreover, existing ABM models of urban location are mostly partial in terms of their handling of the various aspects of urban location choice. Most of these models simulate only residential location choice based on the monocentric assumption of exogenously determined CBD without incorporating job location choice processes—the other critical long-term location decisions made by households and individuals. Others such as ABODE—Agent-Based Model of Origin and Destination Estimation (Tilahun and Levison, 2013) simulate job location choice only while treating the residential location of workers as exogenous. In view of the lack of integration of the residential-job location choice process in these model, it is impossible to match trip production and attraction patterns and to generate origin and destination measures with these models. Thus, the capabilities of new generation of disaggregate urban location choice models would be enhanced significantly by explicitly integrating both residential and job location choice behaviours as the fundamental requirement for estimating the associated patterns of mobility.

Finally, a thorough review of the literature points to the fact that significant amount of both empirical research has been confined to cities in Europe and North American contexts (see e.g. Acheampong and Silva, 2015). It is therefore not coincidence that the various simulation models that have accompanied these studies have also remained essentially confined to these contexts. Improved understanding of the urban location choice process would require broadening the geographical scope of current research. In particular, the existing literature would benefit from new research from the context of cities in the Global South.

1.2 Research objectives and questions

Two main objectives, which reflect both the fundamental empirical questions that need addressing as well as the emerging issues from research trends in simulation model development, highlighted previously, have been addressed in this thesis. The two main objectives and the specific research questions pursued under each of them are outlined as follows:

1.2.1 Empirical research objective and questions

The first objective of this thesis was to:

- examine empirically within a metropolitan context, the location choice behaviour of households and individuals with respect to their residential and job locations, and the mobility patterns associated with the observed home-work location combinations.

To address this objective, four fundamental empirical questions were pursued. These were;

i. What are the factors and underlying processes of residential location choice behaviour of heterogeneous households in urban areas?
ii. What are the factors and underlying processes of job location choice behaviour of individual working members of the heterogeneous urban households?
iii. What are the interdependence between the residential location choice and job location choice of the households? and;
iv. What are the mobility patterns associated with the emergent residential-job location combinations?

1.2.2 Simulation model development objective and questions

The second objective of this thesis was to:

- develop an integrated geospatial and agent-based model to simulate how the residential and job location choice behaviour of heterogeneous households and individuals co-emerge with mobility patterns in a metropolitan area.

The development of the model was driven by the following five research questions:

i. How do the socio-demographic characteristics and preferences of heterogonous households and individuals interact with existing urban structural conditions to influence urban location choice behaviour?
ii. How do bilateral transactions, competitive behaviour and interactions among individual actors in the property market lead to the formation and evolution of property prices?
iii. What are the residential location patterns that emerge from the interaction between households and individuals' choice behaviour and existing urban structural conditions?

iv. What are the employment location patterns that emerge from the interaction between the attributes of individual working members of the households and prevailing job market conditions?

v. How does the emergent residential and job location combinations and individual-level attributes of agents interact to shape home-work mobility patterns?

As would be demonstrated throughout this thesis, the two main research objectives and the specific research questions derived from them are mutually linked. Firstly, the empirical research objective and the accompanying research questions were deployed to anchor the overall research to a specific case study area and to provide a snapshot, static exploration of how households and individuals make their residential and job location decisions as well as the home-work mobility characteristics in the case study context. The empirical studies then led to the development a more dynamic, facsimile[1] simulation model of the co-evolution of urban location choice and mobility patterns within the same metropolitan context. Specifically, the resulting analytical outputs of the empirical research served as inputs into the simulation model development through the formulation and encoding of behavioural rules and heuristics as well as the calibration, verification and validation of the model.

1.3 Overview of the research approach and methodology

While a detailed discussion of the research methodology will be presented in later chapters of the thesis, this section provides an overview of the study design and introduces briefly the case study metropolis selected for this research. Next, an overview of the methodological issues covering the empirical studies and the simulation model development components of the research is presented.

1.3.1 The Case study approach and case study area

The two main objectives of this research, covering the empirical studies and the development of a dynamic simulation model of location choice and mobility patterns, were addressed using a case study design. As underscored by Yin (2013), a case study approach allows for extensive and in-

[1] Facsimile models are those developed to replicate particular phenomenon for a given situation as close as possible (see Gilbert, 2007)

depth analysis and description of a phenomenon within a given physical, socio-cultural, economic and political context. Given that the phenomenon under study—urban location choice behaviour and mobility patterns—has both descriptive and exploratory aspects as reflected in the research questions outlined in the previous sections, the case study design was considered the most appropriate approach for this research.

For this research, the Kumasi Metropolitan Area (KMA) in Ghana, West Africa was selected as a single case study area based on a combination of objective and subjective considerations. As a medium size metropolis covering a total area of $212km^2$ and a resident population of nearly two million, the KMA exhibits many complex dynamics in its land use and transport systems, making it a challenging yet interesting context to study urban location choice behaviour and mobility patterns. The selection of the KMA was also consistent with the quest by this research to contribute to knowledge accretion and expansion by gaining insights from previously unexplored contexts in the Global South. Moreover, located in the researcher's home country, the researcher's background knowledge of the context and existing networks proved very helpful in the overall research design, especially in obtaining the relevant data from both primary and secondary sources to inform this research. Finally, the potential of the research output, particularly, the model development, to benefit the case study area, where the application of urban models as decision support systems remain unexplored despite the rapid pace of urbanization and the associated challenges of urban growth management, also informed the choice of KMA as the case study area for this research.

1.3.2 Overview of the research methodology for empirical studies and model development

Anchoring the first research objective to the Kumasi metropolis, the case study area, a cross-sectional survey was conducted to obtained fine-scale primary data from a random sample of household and individuals regarding their residential and job location choice and the home-work mobility patterns. Details of the survey design including an outline of the study variables, research questionnaire design, sampling techniques and the data analysis themes and statistical methods adopted will be presented later in chapter three.

The second major methodological component of this research, was the specification of a model development framework to achieve the second objective of this research, which was to develop a disaggregate model to simulate the co-emergence of urban location choice and mobility patterns. The formulation of the conceptual model adopted an integrated ABM and Geographical Information Systems (GIS) approach. The framework drew on and integrated under the broader umbrella of complexity theory, the concepts and principles of urban location theory, utility theory, bounded rationality and heuristic decision-making. The framework then provided the conceptual basis for the implementation of the Metropolitan Location and Mobility Patterns Simulator (METLOMP-SIM). Description of the conceptual model using the standard ODD (Overview, Design concept and Details) protocol for ABMs is presented later in chapter six followed by a detailed discussion of the model implementation, verification and validation processes in chapters seven, eight and nine.

1.4 Contribution and innovation

The contributions of this thesis to the literature are demonstrated through the two focus areas covering the empirical analysis of urban location choice and mobility patterns based on cross-sectional data, and the development of METLOMP-SIM, a disaggregate agent-based and geospatial model to simulate dynamically, the co-evolution of urban location choice and patterns of spatial interaction in a metropolitan context.

In chapter four, the first of two empirical studies of the urban location choice and mobility nexus within the context of the Kumasi metropolis is presented. The traditional approach where broad urban-zones constitute discrete choice alternatives in residential location choice models is no longer acceptable as this fails to capture and quantify the extent to which different factors acting across multiple scales influence the choice process. In view of this, using multivariate statistical analysis methods including Principal Component Analysis, Multinomial Logistics Regression and Linear Regression Models, this chapter contributes to a much deeper understanding of how location-defining attributes at the macro, meso and micro spatial scales interact with socio-demographic attributes of heterogeneous households to determine their residential location choice. Maintaining a similar approach, the distribution of individual members of the households' work locations relative to their place of residence, and the determinants of work location choice are also examined. By so doing, a better understanding of the spatial anchors of interaction is provided as the basis to subsequently analyse patterns of spatial flows and the characteristics of these flows. Within this same chapter, the nature of the choice relationship between the place of residence and place of work is analysed. This is done using data elicited on the sequence of choice of the two location pairs initially, the changes that have occurred with these locations over time and how the residential and job locations have responded to each other in cases where there have been changes. This aspect of the study is very important because it establishes the interdependence between the two choice sets and determines whether the simulation of the choice process, which is presented later, is implemented using a conditional choice or co-joint choice assumption.

The analysis of residential and job location preferences and choice interdependence leads to the second empirical study which focuses on quantifying the size of flows between the residential and job location pairs and the attributes of those flows including travel frequency, trip origins and destinations, transport mode choice as well as the associated travel time and costs. Here, the analysis contributes to our understanding of how urban functional structure, which reflect the configuration of employment, residence and the transport networks linking them, provides the spatial anchor for trip production and attraction patterns observed between the home and work locations. In addition, the interplay between spatial variables, socio-demographic attributes of individuals, and attitudes and expectations on transport mode choice is presented using a series of logistic regression models. In particular, the determinants of choice of different work travel modes such as between motorized and non-motorized (i.e. walking) transport, between private and public transport and between different public transport modes are established. Besides mode choice, this chapter also makes important contribution to our understanding of travel times and costs across different groups of commuters in the case study metropolis.

Perhaps, the most significant contribution of this thesis is development of METLOMP-SIM, a disaggregate model that simulates the co-emergence of urban spatial structure and mobility patterns as a function of the interaction between individuals' location choice behaviour in the urban property and job markets, and existing urban structural conditions. One of the major contributions of the model development is that while the empirical studies provided a snapshot understanding of the relationship between the long-term urban location choice decisions and short-term choices related to daily patterns of mobility, METLOMP-SIM, being a facsimile model, builds on the empirical insights derived to simulate dynamically, the co-evolution of location choice and mobility patterns in the metropolitan context. Using an innovative agent-based, computer modelling paradigm, a systematic approach of specifying a conceptual model based on established conceptual and theoretical principles, linking the framework to data and encoding agents' rules of behaviour and heuristics from the bottom-up to generate urban location and mobility patterns observable at different spatial scales is presented in METLOMP-SIM.

Moreover, while building on the contributions of previous modelling attempts, the current model presented in this thesis distinguishes itself from its predecessors in several ways. First, METLOMP-SIM extends existing models in the ways in which various components have been represented to improve model realism–the extent to which computer models represent the real-world systems they seek to replicate. The explicit integration of property market dynamics including bilateral transactions and competition among agents which account endogenously for price formation brings METLOMP-SIM much closer to reality than most of the models that have preceded it. Secondly, the representation of both macro and meso level locational attributes as well as specific spatial goods (i.e. land and dwellings) as differentiated discrete choice alternatives, makes significant improvement with this model. Contrary to existing models that tend to focus primarily on residential location choice in a buyers' market, this model simulates all types of tenancies—market and non-market—to reflect as close as possible, the real-world conditions within which the population exercise choice. Moreover, METLOMP-SIM distinguishes itself from predecessor models of location choice by explicitly simulating job location choice. With job search process integrated in the model, METLOMP-SIM provides additional capabilities by making it possible to derive and analyse the mobility patterns associated with emergent residential and job location combinations such as trip origins and destinations, distance, and mode choice.

1.5 Thesis structure and organization

Besides this introductory chapter, which has set out the overall research context and outlined the research objectives, questions, methodology and contributions to the literature, the rest of the thesis is organized into nine chapters.

In chapter two, an account of previous research relevant to this thesis is presented. The literature review discusses the dominant theories and concepts as well as important empirical studies that have accumulated over the years. It also discusses the dominant urban modelling traditions and the various types of models that have emerged from these traditions. By tracing important developments and the general direction of research in the literature, gaps are identified based on which the key objectives and questions addressed in this thesis are formulated.

In chapter three, the first of two research methodologies deployed to address the objectives and questions of this thesis is presented. This chapter discusses the case study design adopted in this research. It anchors the overall research to a specific metropolitan context and proceeds to discuss important methodological issues including the survey design, sampling techniques, data collection as well as statistical methods adopted to analyse the data.

The fourth chapter constitutes the first of two data analysis results chapters presented in this thesis. It focuses on the analysis of the urban location choice process of households and individuals, examining the residential-job location preferences and the interdependence between the two choice sets based data obtained through a cross-sectional survey from the case study metropolis. Thus, this chapter addresses the first three research questions outlined under the empirical objective of this thesis.

In chapter five, the second data analysis results focusing on the home-work mobility patterns observed in the case study metropolis is presented. Here, various dimensions of flows and interactions between the home and work location pairs of individuals, including home-work trip frequency, trip origins and destinations, travel mode choice, travel distance, travel time and commuting costs, and their determinants are examined.

The sixth chapter initiates the steps towards addressing the second objective of this thesis which was to develop a disaggregate model to simulate how urban location choice co-evolves with mobility patterns. In this chapter, a conceptual model is specified drawing on the relevant theories and concepts. The overall structure of the model, indicating the model sub-components, entities, parameters and variables as well as the ABM design concepts principles underpinning the model and the decision-making framework of agents in the model are discussed.

The seventh chapter focuses on the implementation of the integrated geospatial and agent-based model of residential location choice and mobility patterns. It presents the simulation platform and programming language adopted for the model development. Details of the model initialization and execution processes are also provided.

In chapter eight, the calibration of the model and analysis of the simulation results are also presented in this chapter. This chapter focuses on the series of model calibration and parameter sweeping experiments conducted as well as the model simulation results in terms of the simulated residential location patterns and property market price formation and evolution, the simulated job location patterns and the simulated home-work mobility patterns

Chapter nine addresses an important aspect of the model development process which is verification and validation. Discussed here are the processes followed to ensure that the model has been correctly implemented and that it corresponds to and explains urban location choice and mobility patterns in the case study metropolis.

The tenth chapter discusses the main finding of this thesis. It relates the findings of this research to the wider literature and critically evaluates its relative strengths and limitations. The chapter also discusses the possible applications of this research and concludes by offering possible areas for future research.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

Land-use configuration and mobility patterns in urban areas co-evolve. The long-term location decisions made by individuals, households, firms and public sector institutions determine the distribution of land use activities such as the place of residence, employment and ancillary facilities across the urban landscape. The distance separation between the land use activities, in turn, determines the commuting possibilities and daily patterns of human spatial interaction.

This fundamental urban structure and travel demand co-determination conception was first examined empirically by Hansen (1959) in Washington DC. Over the past sixty years, significant research has focused on understanding and disentangling the mechanisms underlying the co-emergence of urban structure and travel demand. Drawing on theories and concepts in urban economics, discrete choice modelling, social physics, systems analysis and the complexity sciences, research in the field has focused on two broad, mutually dependent areas. Firstly, significant research has focused on gaining empirical insights into how urban structural characteristics impact patterns of spatial interaction and vice versa. The second broad research focus area has been on the development of state-of-the-art urban land use and transport simulation models for both exploratory research and practical policy application purposes.

The aim of this chapter is to review the existing literature on the urban structure and mobility nexus. In doing so, the dominant concepts and theories underpinning research in the field are discussed. The chapter also discusses the accumulated empirical evidence and evaluates the different methodological approaches that have been adopted in the field to model the relationship between the land use and transportation systems. The rationale is to identify areas where new contributions could be made in terms of both empirical investigations and simulation model development, and to identify the appropriate methodologies for addressing the research gaps.

2.2 Chapter organization

The remainder of the literature review is structured into six sections. A review of the conceptual and theoretical foundations of the relationship between urban spatial structure and travel is first presented. This is followed with a discussion of empirical studies on the topic focusing on the location choice and travel demand literature. Next, existing models of urban land use and mobility patterns are evaluated focusing on modelling paradigms and the accompanying methodological traditions that have been adopted. A critical evaluation of current research directions and gaps, in both the areas of empirical research and model development is then presented. Based on the gaps identified, the research objectives and questions to be addressed in this thesis are formulated in the penultimate section. The chapter ends with a summary of the literature review and opens a window into the next chapter.

2.3 Understanding urban land use and transport nexus: concepts and theories

In broad sense, the term land use, comprises various sub-systems such as residential, industrial, commercial and physical infrastructure, which reflect the location of human activities such as residence, work, shopping, schools and recreation (Mackett, 1993). The transport system encompasses the network of flows as well as the physical infrastructure and the supporting services which provide accessibility opportunities between spatially segregated activity locations (Wegener and Fuerst, 2004). Both systems are the outcome of the interplay of complex bottom-up and top-down processes generated by several urban actors such as households and firms making location choice decisions and public sector institutions' formal policy responses in the domains of urban land use planning and transportation planning.

The above components and functions of the land use and transport systems imply that the two systems are linked. The 'land-use transport feedback cycle' (Wegener, 2004) conceptualizes the nature of the link between the two systems as a complex, two-way dynamic relationship. According to this framework, the distribution of land use determines the location of activities. The need for interaction arises as a consequence of the spatial separation between the land-use activities. The transport system creates opportunities for interaction or mobility, which can be measured as accessibility, and the distribution of accessibility in space, over time, co-determines location decisions and results in changes in the land use system. Thus, the key concept that links the two systems is accessibility, defined broadly as the ability to reach activity sites distributed in space within a given period (Hanson and Giuliano, 2004).

In addition to the above conceptual framework, several theories from different disciplines including urban economics, urban geography, social-physics and the complexity sciences have been advanced to describe and explore the processes underlying the configuration of land use and travel demand. The dominant theories in the field can be grouped under five main categories. These are; aggregate urban economics and gravity-based entropy maximization theories; random utility theory; theories of decision-making under uncertainty; time-geography theory and general systems and complexity theory. In the next section, an overview of the theories is presented.

2.3.1 Aggregate urban economics and gravity-based entropy maximization theories

Classical urban micro-economic models of Alonso (1964), Ricardo (1821), Von Thunen (1826) and Wingo (1961), posit that transport cost, a function of travel distance, has profound impact on the location of activities and the overall optimum emergent structure of cities. These access-space-trade-off models, grounded in urban economics theory, rely on simplifying assumptions including monocentricity, spatial homogeneity, rationality in decision making and systems in equilibrium to provide qualitative analysis of the relationship between location and transport. Consequently, these models are deterministic in nature, restrictive in their application, and are unable to capture the richness of urban and regional geography required to fully represent the complexities in the land use and transport systems (de la Barra, 1989, Waddell, 1997).

In the early 1960s, the theory of social physics, grounded in the Newtonian concept of gravity gained recognition as a practical approach to model mathematically, the relationship between urban structure and patterns of spatial interaction. Popularized by Lowry (1964) in his Model of the Metropolis developed for the city of Pittsburgh, the gravity concept together with empirical analysis of human spatial interaction behaviour became the theoretical basis for a new generation of models, called Spatial Interaction Models.

The basic Lowry gravity model states that the interaction between any two zones is proportional to the number of activities in each zone and inversely proportional to the friction impeding movement between them. Although initially adopted due to its conceptual simplicity and mathematical tractability, it was not until after close to a decade that Wilson (1970) drew on the concept of entropy maximization to provide a general theoretical foundation for the gravity approach. Entropy refers to the degree of disorder in a system, which in the context of land use and transport systems, results from the relative location of workers, jobs and housing in the city (de la Barra, 1989). Within the framework of entropy maximization, the amount of interaction between activity zones can be worked out as a doubly constrained, origin-constrained, destination-constrained, or an unconstrained matrix model (Batty, 2013).

2.3.2 Random utility theory

McFadden's (1973) Random Utility Theory (RUT) gained prominence in land use and transport interaction research out of the quest for a robust framework that could capture the complex choice behaviour decisions at the individual level while overcoming the weak assumptions and misspecification errors inherent in aggregate spatial interaction and urban economic approaches.

According to McFadden's utility theory, choice between alternatives are predicted as a function of attributes of the alternatives, subject to probabilistic variations in the knowledge, perceptions, taste, preferences, and socio-economic characteristics *inter alia* of decision makers. Utility theory therefore allows one to develop location choice and travel behaviour models based on the study of disaggregate behaviour (Iacono et al., 2008).

Contrary to gravity-based models, utility-based models can effectively address locational characteristics using a bundle of locational attributes, with each element in the bundle reflecting a distinct feature of the location, and a random component representing the unobserved characteristics of a location (Chang, 2006). Despite enjoying sound theoretical foundation, utility-based models have been criticized for their inability to explicitly capture the underlying decision processes and behavioural mechanisms that result in observed location-travel decisions (Ettema, 1996, Fox, 1995, Pinjari and Bhat, 2011).

2.3.3 Theories of decision-making under uncertainty and bounded rationality

Classical utility theory presented in the previous section assumes rationality and perfect information in choice decisions. However, within the transportation and activity systems, decision makers face conditions of uncertainty, for example, in choosing departure times, activities, destinations, transport modes and routes (Rasouli and Timmermans, 2014a). Theories focusing on decision making under uncertainty offer a framework to relax these unrealistic assumptions and limitations. Decision making under uncertainty is viewed as a choice between gambles or lotteries (Tversky, 1975). Within this framework, the characterization of the choice alternatives is captured in terms of probability distributions where individuals are not sure about the exact state of the choice alternative or about the outcome of their decisions (Rasouli and Timmermans, 2014a).

A survey through the literature shows three standard theories of decision making under conditions of uncertainty being applied to transportation research. These are expected utility theory (Bernoulli, 1738; von Neumann and Morgenstern, 1944; Savage, 1954), prospect theory (Kahneman and Tversky, 1979) and regret theory (Bell, 1982; Fishburn 1989; Loomes and Sugden, 1987).

Expected utility theory (EUT) was formulated as a descriptive model of economic behaviour. The foundational contribution is linked to the so-called St. Petersburg paradox—the puzzle surrounding what price a reasonable person should be prepared to pay to enter a gamble, a game of infinite mathematical expectation, consisting of flipping a coin as many times as is necessary to obtain 'heads' for the first time (Bernoulli, 1738). EUT states that the decision maker chooses between risky or uncertain prospects by comparing his or her expected utility values—the weighted sums obtained by adding the utility values of outcomes multiplied by their respective probabilities (Mongin, 1997). Critical evaluation of the limitations of EUT and efforts devoted toward developing alternatives to EUT can be found in Starmer (2000) and Kahneman and Tversky (1979).

An extension of EUT is prospect theory (PT) (Kahneman and Tversky, 1979). The key principle underpinning the theory is that decisions are made based on the potential value of loss and gains rather than the final outcomes. (PT) posits that these losses and gains are evaluated using heuristics. Proponents posit a two-stage decision-making process. The first stage involves the use of various decisions to frame possible outcomes in terms of gains and losses, relative to some neutral reference point, while the second stage involves evaluation of the outcomes of each alternative according to some value function, which transforms objective probabilities into subjective probabilities (Rasouli and Timmermans, 2014a).

Closely related to PT is bounded rationality theory (Simon 1957, 2000; Tversky, 1969). Taking their roots from social psychology and behavioural economics, proponents argue that decisions are made on subsets of factors, affected by perpetual cognitive biases, uncertainty and information asymmetry, and do not necessarily result in optimal choices (Payne, et al., 1993; Innocenti et al., 2013; Zhu and Timmermans, 2010). Within this context, heuristics—simple, quicker and practical rules—are used to guide decisions when faced with uncertainty. Leong and Hensher (2012) in their review, identified four types of heuristics strategies employed by individuals in their choice behaviour namely; satisficing, lexicography, elimination-by-aspects, and majority of confirming dimensions.

Regret theory (RT) is attributed to seminal works of Bell (1982), Fishburn (1989) and Loomes and Sugden (1982, 1987). The theory is grounded in "the notion that individuals' utility of choosing an alternative is not only based on the anticipated payoff of each individual choice alternative across different states of the world, but also on anticipated payoff of the other alternative" (Rasouli and Timmermans, 2014a, p8). Thus, RT focuses on the opportunity loss in decision making—the difference between actual payoff and the payoff that would have been obtained if a different course of action had been chosen.

2.3.4 Time-geography theory

An equally important theoretical tradition relevant for land use and transport research is the time-geography paradigm attributed to the original work of Hagerstrand (1970) and Chapin (1974). The time-geography paradigm posits that spatial interaction occurs within a framework of spatio-temporal constraints, which necessitates trading of time for space (Miller and Bridwell, 2009; Peters et al., 2010). Conceptually, time-geography theory uses a space-time prism to analyse the envelope of possibilities open to an individual, subject to several spatio-temporal constraints

(McNally, 2000). Crease and Reichenbacher (2013) and Miller (2005) identified three main spatio-temporal constraints of spatial interaction namely: *capability constraints*, the ability or otherwise of an individual to overcome space in time; *coupling constraints*, arising from the need to undertake certain activities with other people for given durations; and *authority constraints*, resulting from common social, political, cultural and legal rules as well as exclusionary mechanisms that restricts an individual's physical presence at a location. As will be discussed later in this chapter, the time geography paradigm has over the past two decades, influenced the development of activity-based travel demand models.

2.3.5 General systems and complexity theories

Complexity theory and general systems theory (von Bertalanffy, 1950; Boulding, 1956; Forrester 1993) have also gained recognition in the field of urban and regional planning in general. As a contemporary embodiment of general systems theory (Batty, 2007; 2013), complexity theory provides the framework to think about cities as complex adaptive systems with several interacting components that manifest perpetual disequilibrium (Albeverio, 2008; Batty, 2017). Applied in the context of land use and transport modelling, complexity theory provides a robust framework to study the path-dependent and emergent behavioural outcomes of urban actors as well as the dynamic feedback relationship between the land-use and transportation systems. On-going efforts to develop computer simulation models, using micro-simulation approaches to capture complex interactions of linked responses that lead to a co-evolution of urban structure with mobility patterns, are grounded in systems and complexity theory (Albeverio, 2008; Batty, 2007; Samet, 2013).

In summary, research over the past six decades has drawn on several theories that can be applied either at an aggregate or disaggregate level of understanding decision-making behaviour within the land use and transport system. Urban economics theory and entropy-based gravity models allow for macro-level analysis using simple and tractable mathematical formulations. Classical utility theory and theories of decision making under uncertainty both focus on the complex micro/individual-level choice between discrete alternatives. The time geography paradigm, allows one to quantify activity-travel patterns within a framework that explicitly incorporates spatio-temporal opportunities and constraints. Complexity theory, provides an umbrella framework under which various theories and methodologies could be integrated to explore system interdependence and emergence based on a bottom-up thinking of the interaction of several actors and their environment.

2.4 Empirical studies of the relationship between urban structure and mobility patterns

The concepts and theories discussed in the forgoing sections have inspired significant empirical enquiry in the field of land use and travel behaviour studies. These studies have sought to disentangle the mechanisms through which the complex interaction among several forces of physical, socio-demographic, economic and policy changes give rise to the observed patterns and behaviour within the urban land use and transport systems in space and time.

Investigating the causal mechanisms by which urban form or urban built environment characteristics influence travel behaviour and vice versa, constitutes one of the dominant focus areas of empirical research. These studies integrate built-environment/urban structural characteristics and behavioural attributes of households and individuals. The built environment characteristics include density, land use mix, home-employment balance, neighbourhood design, street network connectivity, destination accessibility and distance to transit. The behavioural component addresses households and individuals' characteristics such as income, gender, professional status, life-cycle stage, attitudes, beliefs and perceptions.

Based on cross-sectional data, and controlling for several socio-demographic variables in multivariate regression models, some studies have concluded that urban structural variables have significant influence on travel behaviour (e.g. Næss, 2013; Gim, 2013; Aditjandra et al, 2013; Eboli et al, 2012; Nam et al., 2012; Grunfelder and Nielsen 2012; Ewing and Cervero, 2010). Others have adopted longitudinal and quasi-longitudinal design, incorporating cognitive constructs such as attitudes, preferences and perception as mediating factors to establish a strong causal link between built environment attributes and travel (Handy et al, 2005; Meurs and Haaijer, 2001). Thus, these studies show that urban structure determines the locations where different activities may take place and combine with personal characteristics to determine different kinds of travel behaviour. For example, populations living in sub-urban areas, where densities tend to be relatively low, tend to travel longer distances using private cars compared to inner-city residents among whom active transport (walking or biking) constitute the main mode of transport.

Besides investigating the influence of urban form on travel behaviour, other empirical studies have focused on understanding location choice behaviour of households and individuals and the associated mobility outcomes. In these studies, the long-term location choice behaviour of households and individuals regarding where to live and to work constitute the primary focus. The emphasis on residential and job location choice derives from the recognition that these choice decisions have long-term consequences (Pinjari and Bhat, 2011). Firstly, households' residential locations and individuals' job locations once made, remain stable over long periods of time. Secondly, the location where people live and work, has profound impacts on the form and functional structure of urban areas, as well as the patterns of spatial interaction (Yang and Ferreira, 2008). Indeed, residential and employment activities constitute the two major land uses in urban areas.

Existing studies have established that residential location choice is influenced by several factors comprising location-defining attributes at the urban and neighbourhood scales as well as dwelling-defining attributes at the micro-scale. Specifically, factors such as noise levels, municipal taxes, housing prices and rent levels (Hunt, 2010; Pagliara et al., 2010), available mobility options including proximity to bus lines (Boschmann, 2011), density of development and access and to high-quality schools (Pagliara et al., 2010) influence people's residential location preferences. Other determinants of residential location choice identified in the literature are the degree of commercial or mixed land-uses in an area, incomes and neighbourhood composition (Pinjari and Bhat 2011), social networks (Tilahun and Levinson 2013; Acheampong, 2016), life-stage and the evolution of household membership and family structures over time (Habib et al., 2011, Lee and Waddell, 2010).

Similarly, employment location choice is influenced by a range of factors including opportunities for higher wages, home-work distance separation, job locations of all household members, proximity to essential amenities and the education and skills match between available jobs and job-seekers (Tilahun and Levison 2013, Kim et al., 2001; Moeckel, 2016; Glaeser and Resseger, 2010).

Furthermore, the relationship between residential and job location choice processes has gained attention in the literature. Early research, drawing on the access-space-trade-off framework derived from classical urban economics theory assumed that workplace choice is predetermined or exogenous to residential location choice (Waddell et al. 2007). More recent empirical works (e.g., Boschmann 2011, Habib et al., 2001, Kim et al., 2005, Pinjari and Bhat 2011, Waddell et al. 2007, Yang et al., 2013; Inoa et al, 2015) have, however, established that initial residential and job location choices as well as subsequent housing and job mobility decisions are jointly determined. The implications of the choice interdependence between the two location combinations for further empirical studies and location choice modelling will be one of the focus under the discussion of research directions and gaps presented later in section 2.7 of this chapter.

2.5 Modelling urban land use and travel demand: aggregate vs disaggregate modelling approaches

Aside from empirical observations, a considerable amount of research efforts has gone into developing models of urban land use and travel demand. These models have either fundamental exploratory research purposes or decision-support purposes such as being used to assess the long-term impacts of land use and transport policies on the distribution of activities and travel demand.

Different modelling paradigms inspired by the various theories presented earlier in this chapter have been adopted to develop both prototype and operational models. The literature of urban modelling reveals two main approaches to modelling urban land use and travel demand. These are aggregate modelling approaches grounded in traditional urban economics and entropy-based gravity traditions and disaggregate approaches which focus on building models that explicitly represents the actions and behaviour of individual agents located in space (Batty, 2016). Two main disaggregate modelling approaches can be identified in the literature. These are utility-based econometric modelling, and micro-simulation approaches to which techniques such cellular automata and agent-based modelling belong. In the sections that follow, these modelling traditions are briefly discussed and followed with a discussion of models that have been developed based on them.

2.5.1 Aggregate urban economics and entropy-based gravity modelling approaches

Under the aggregate modelling approach, both space and activities are grouped into discrete categories of zones containing larger number of activities while individuals are put into groups assumed to share similar characteristics. These aggregate zones interact to generate flows in the form of trips which are also aggregated at the level of urban-zones. The aggregate modelling tradition therefore draws on the bid-rent and access-space-trade-off formulations of Alonso (1964) Wingo (1961) and entropy-based gravity modelling tradition popularized by Lowry (1964) and Wilson (1970). Models adopting this approach also maintain the simplifying assumptions of urban micro-economics theory such as mono-centricity, spatial homogeneity and rational and homogenous urban actors. As such, the aggregate modelling approach does not allow one to represent spatial heterogeneity, important urban market dynamics as well as the behavioural aspects of heterogeneous decision-makers that given rise to activity distributions and travel demand (Shirmer et al., 2014; Bhat, 2015).

2.5.2 Disaggregate utility-based econometric modelling approach

The utility-based econometric modelling approach derives theoretical foundations from McFadden's random utility and discrete choice theory. The fundamental assumption underlying this approach is that individuals are rational decision-makers whose aim is to maximize utility relative to their choices. This approach allows the study of disaggregate behaviour of individuals (e.g. housing choice and mobility choice) by representing the attributes of discrete alternatives such as dwellings differentiated by intrinsic characteristics and travel mode. Utility-based econometric models are often systems of equations that capture relationships between individual-level socio-demographics and activity-travel environment to predict probabilities of decision outcomes (Pinjari and Bhat, 2011). As a disaggregate approach, it provides a framework to model choice as a function of a finite number of discrete alternatives, subject to variations in attributes of the decision-maker such as knowledge, taste, preferences, attitudes, income, among others (Chang 2006, Pinjari and Bhat 2011; Cascetta, 2009).

2.5.3 Disaggregate micro-simulation modelling approaches

In recent years, the application of micro-simulation approaches in transportation and land-use research have gained prominence among researchers in their bid to replace aggregate probabilistic modelling approaches with disaggregate stochastic approaches. The disaggregate approach is therefore popular among models built around representing the actions and behaviour of individual agents located in space (Batty, 2016). These disaggregate approaches derive theoretical inspiration from urban systems and complexity theory. The micro-simulation tradition to which techniques such as cellular automata (CA) and agent-based modelling (ABM) belong, involve the simulation of the aggregate behaviour of a system as the sum of the actions and interactions of disaggregate behavioural units within the system (Iacono et al., 2008; Miller and Savini, 1998).

A micro-simulation system could be a procedural program in which data flows are computed and driven by statistics and probabilities algorithms (Hunt et al. 2008, Wagner and Wegener 2007; Waddell, 2001). The approach uses computer simulation programs/algorithms of rules in the form of condition-action (if-then) pairs to mimic individual decision-making processes (Batty, 2014). In simulating choice processes, micro-simulation often uses Monte Carlo simulation methods where random numbers are used in the process of deciding which of the available alternatives the decision-maker will choose, given the calculated probabilities (Feldman et al., 2010).

The ABM paradigm as a micro-simulation approach, takes a bottom-up perspective for the creation, analysis and experimentation with urban models composed of autonomous agents that interact with each other and their environment locally (Gilbert, 2008; Railsback and Grimm, 2011; Railsback, et al., 2006; Wu and Silva, 2010). ABM as a modelling technique allows for a natural description of a complex system in a flexible and robust manner to capture emergent phenomenon (Batty, 2001; Bonabeau 2002; Castle and Crooks, 2006; Wu and Silva, 2010; Silva, 2011). Given that agents' behaviour can be driven by adaptation and evolutionary learning, ABMs become suitable in exploring new dynamics and critical transitions in complex systems (Huang et al., 2014). Moreover, the ABM approach allows the agents (e.g., household members or individuals within the simulated population) to learn, modify, and improve their interactions with their environment (Batty, 2007; Jin and White, 2012; Silva, 2011). The ABM paradigm also allows one to integrate both inductive and deductive approaches to simulate emergence of aggregate patterns in complex systems (Nolan et al., 2009).

Disaggregate micro-simulation modelling approaches have several advantages over the aggregate approaches. While micro-simulation/ABM and utility-based methods tend to be disaggregate models, the main advantage of the former over the latter is that it allows one to model the increasing heterogeneity of the urban lifestyle, new tendencies in mobility behaviour as well as environmental impacts of land-use and transport policies at the necessary spatial resolution (Hunt et al., 2008; Wagner and Wegener, 2007). As a disaggregate modelling approach, micro-simulations derive strength from their dynamic nature, which makes it possible to trace model components (e.g., individuals, households, jobs, and dwellings) over time to observe the modelled processes of change at a level of detail that is not possible in other types of models (Pagliara and Wilson, 2010).

Moreover, these disaggregate modelling approaches allow for the integration of several theoretical principles from random utility theory, bounded rationality, choice heuristics and systems theory into a composite framework to simulate complex urban systems. Finally, a disaggregate approach allows for the integration of a wide range of data sources including expert knowledge, empirical research, insights from existing models and survey data as a context to formulate agent behaviour (Murray-Rust et al., 2013; Janssen and Ostrom, 2006; Robinson et al., 2007).

2.6 Overview of models of land use and travel demand

Several land use and travel demand simulation models have been developed over the years. These models integrate different theoretical principles, empirical evidence and modelling approaches to represent various aspects of the land use and transport system either separately or as integrated models. It is worth mentioning that, although the fundamental relationship between land use and transportation is widely recognized, the development of models in the two fields have traditionally evolved separately, with some attempts to integrate the two systems. Consequently, within the context of this research, three main types of models according to purposes can be identified. These are:

i. Urban land-use models developed to simulate location choice (i.e. residential and job location choice);
ii. Urban models that focus exclusively on modelling travel decisions as the basis to understand and predict travel demand; and
iii. Land-use and transport interaction (LUTI) models that integrate both location choice decisions and travel decisions into a composite framework.

An overview of the different types of models under these three categories is provided in the following sections emphasizing model components, the underlying theoretical principles as well as the accompanying modelling approaches.

2.6.1 Land use models of urban location choice: Emphasis on disaggregate models

Urban land use models of location choice are models that focus on understanding the emergence of spatial structure as a consequence of the location choice decisions of urban actors. The central focus of these models often is to simulate the residential (re)location choice behaviour of households and individuals. Initially, residential location choice was modelled at aggregate levels based on the bid-rent and access-space-trade-off framework. Within these models, households of similar characteristics would choose between different urban-zoness of some location- defining and activity characteristics. Households were assumed to a make trade-off decision between locations closer to a central location which serves as the focal point of all major urban functions to reduce transport costs or locations farther away from the central location where land is available at relatively cheaper price.

In recent years, the development of residential location choice models has adopted a more disaggregate approach. These models combine economic theory with micro-simulation techniques to simulate the choice behaviour of heterogeneous agents. These models have been evaluated here based on the level of detail they capture with respect to agent heterogeneity, types and attributes of discrete choice alternatives and representation of bilateral transactions and competition in the property market. These criteria are important because they determine the extent to which these models can realistically represent the real-world phenomenon they seek to replicate.

Table 2.1. provides a summary of the existing disaggregate models of residential location choice which have been developed using agent-based modelling approaches per the above criteria. Firstly, disaggregate residential location models are a complete departure from aggregate access-space-trade-off models in that they represent heterogeneous sets of agents. The number of different types of decision-makers as well as basis of differentiation however, differ among the various models. Some models define different classes of households based solely on life-cycle-stages (e.g. Fontaine, 2010; Haase et al, 2010). These models are grounded in empirical insights that show that life-course events and changes in individual or household circumstance such aging, marriage, child-bearing, retirement and career changes, among others, influence initial residential location decisions and subsequent decisions to move or change residence.

Table 2.1 Components of disaggregate models of residential location choice

Model developers/ Authors	**Model component and level of details**			
	Agent Heterogeneity	Discrete choice alternatives		Bilateral Transactions
		Land parcels	Dwellings	
Ettema (2011)	+		+	+
Fontaine (2010)	+		+	
Filatova et al, (2011)	+	+		+
Filatova et al., (2009)	+	+		+
Magliocca et al (2011)	+	+	+	+
Hosseinali et al., (2013)	+	+		
Li and Liu, (2007)	+	+		
Tannier et al., 2015	+		+	
Murray-Rust et al., (2013)	+	+		
Robinson and Brown, (2009)	+	+		
Brown and Robinson, (2006)	+	+		
Otter et al., (2001)	+	+		
Lemoy et al. (2010)	+	+		+
Torrens (2007)	+	+		
Parker and Filatova, (2008)	+	+		+
Jjumba and Dragićević (2012)	+	+		
Loibl and Toetzer (2003)	+	+		
Brown et al., (2008)	+	+		
Haase et al. (2010; 2012)	+		+	
Zhuge et al. (2016)	+		+	+

Other models have adopted income as the single criteria to differentiate households (e.g. Lemoy et al, 2010; Hosseinali et al., 2013). In these models, households are grouped according to earnings such low, medium or high income levels. The third differentiation criteria combine both the life-cycle-stage approach with income classification to define different types of household agents. This hybrid approach increases model complexity without compromising realism, and therefore allows one to generate reasonably realistic model outputs based on a wide range of factors that determine the defining attributes of actors. For example, MobiSim—an individual-based model of residential choice developed by Tannier and Colleagues (2015), represent some 72 different types of households based on different combinations of variables including income, age, household composition and number of children.

Secondly, the extent to which dynamics of the property are represented in these models also differ considerably. Some models, as shown in table 2.1, do not explicitly model bilateral property market transactions and associated competition involving bidding, counter-bidding and price formation (e.g. Hosseinali et al., 2013; Murray-Rust et al., 2013). These models often consider the supply side of the property market (i.e. land or dwelling units) as exogenous and focus only on the demand side decision-making processes. Even so, they do not represent the micro-economic decision interaction of agents (i.e. buyers and sellers) that results in price formation and different market power scenarios.

In the model of Hosseinali and Colleagues (2013) for example, they simulate urban development scenarios as a consequence of the decision of heterogonous household agents searching for suitable locations and developing land for housing, based on some measures of accessibility, attractiveness and land values. The only form of market competition represented in their model is when some cells (i.e. land parcels) are simultaneously chosen for development by more than one household agent. Competition outcomes for any given land in their model, depends on the number of times that an agent has been involved in and lost in previous competitions rather than on market interactions among household agents expressing willingness to pay and subsequently engaging in a bidding process based on asked and take price evolving from the transactions. A similar approach where land conversion events are not based on micro-economic decision-making is adopted by Robinson and Brown (2009). In their model, farm and residential sales occur probabilistically based on land lot characteristics rather than on the basis of market principles in which competing land uses can be valued, and the economic constraint or opportunity costs of the acting agents brought to bear in their decision-making processes.

As Ettema (2011), Magliocca et al. (2011) and Filatova et al, (2009) note, the lack of bilateral market transaction and competition in these models makes it difficult to gain insights into the underlying economic forces driving land development decisions. This limits the realism of such models as decision support systems.

In recent years, new generation of agent-based models of location choice and land use decisions have sought to explicitly incorporate market processes involving bidding between buyers and sellers and price formation outcomes (see table 2.1). These models explicitly represent market outcomes derived from the bilateral transactions between buyers and sellers' expectations and perceptions of market opportunities. They specify buyers' willingness to pay (WTP) and sellers' willingness to accept (WTA) respectively. WTP and WTA are then adjusted to form bid and ask prices, thereby accounting for different market power scenarios.

In Ettema's (2011) model of residential relocation for example, house prices emerge through bilateral transactions between buyers and sellers. Transactions are constrained by agents' budgets, housing preferences and their perceptions of market conditions. CHALSM (Magliocca et al., 2011) simulates development density patterns through coupled housing and land-market models linked through supply and demand functions of the developer and consumer households respectively. Mechanisms of land and housing market transactions in CHALSM are built on bilateral transaction framework that links the developer's rent expectations in the housing market to his bid price in the land market. Adaptive expectations of future prices and market conditions are used to compare the utility of present and potential future transactions. Parker and Filatova's, (2008) location choice model include direct modelling of price formation and market transactions that leads to the emergence of land rent patterns without the restrictive assumption to identify prices in equilibrium.

Despite the advances made towards incorporating micro-level bilateral transactions and price formation in new generation of agent-based location choice models, there are several issues that require addressing in these models as far as the representation of the spatial goods (i.e. land or dwelling units) as discrete choice alternatives are concerned. Some models focus on unique parcels of land lots as the only discrete alternatives of households' choice decisions (see table 2.1).

A handful of residential location choice models represent dwellings as discrete choice alternatives. Ettema (2011), in his model, define dwellings as spatially fixed agents. However, the model does not specify the distinguishing characteristics of dwellings such as type, size and type of tenancy it is meant for (i.e. whether for rent or for owner-occupation). Another drawback of Ettema's model is that it simulates a market of relocating households given a fixed dwelling stock and individual dynamics of households only wanting to buy or sell residential properties. Zhuge et al. (2016) also models only owner-occupier decision where agents are on the market to buy a property; renters who do not have a house are not considered in the model. In the model developed by Magliocca et al (2011), they distinguish between eighteen different types of houses determined by different combinations of three different house sizes and six different lot sizes: housing type is defined by lot and housing size, noted as either small, medium or large house. However, the model does not provide detailed categorization into typology (i.e. whether flats, semi-detached or detached dwelling units) and their attributes, (e.g. number of bed rooms).

Although the residential location choice models evaluated here make significant attempts to address the drawbacks associated with initial aggregate models, none of these models explicitly models job location choice. In most of these models, the traditional assumption of exogenously determined employment centre (in the case of mono-centric models) or centres (in the case of polycentric models) is maintained without necessarily simulating the job search processes of individuals as part of the residential location choice process. Some attempts have been made to simulate employment location choice. Tilahun and Levison (2013) have developed ABODE—Agent-Based Model of Origin and Destination Estimation in which they explicitly simulate job search and marching processes arising from of the interaction between firms and individual job seekers. In ABODE, worker agents look for employment opportunities represented as firms distributed in the city, weigh offers and decide positions based on wages offered and their skills. The model however, does not explicitly simulate residential location choice. Instead, it takes the residential location of workers and the locations of employers as exogenous.

2.6.2 Travel demand models

Travel demand models focus on understanding travel behaviour as a basis for predicting and managing travel demand. The key issues of concern in these models, therefore, include trip origin and destination, transport mode choice, vehicle ownership, and trip scheduling/sequencing behaviour. Two main approaches to modelling travel demand can be found in the literature. These are the four-step, trip-based travel demand modelling approach and the activity-based modelling approach. The key features, strengths and limitations of these two approaches are discussed in the sections that follow

i. The four-step transport demand model

Gaining prominence from the 1950s, the four-step travel demand model (FSM) has become the traditional tool for forecasting demand and evaluating performance of transportation systems and large-scale transport infrastructure projects (McNally, 2000). Travel is modelled using trips as the unit of analysis based on origin-destination (O-D) survey. The spatial unit within which trips occur is represented as several aggregate Traffic Analysis Zones (TAZ) defined based on socio-economic, demographic, and land-use characteristics (Bhat and Koppelman, 1999; Fox, 1995; Martinez et al., 2007).

The typical FSM consists of four distinct steps of trip generation, trip distribution, modal split, and route assignment. Each step is intended to capture intuitively reasonable questions relating to: how many travel movements are made, where they will go, by what mode the travel will be carried out, and what route will be taken based on aggregate cross-sectional data (Bates, 2000).

Trip generation measures the frequency of trips based on trip ends of production and attraction to estimate the propensity and magnitude of travel. At the *trip distribution* stage, trip productions are distributed to match the trip attractions and to reflect underlying travel impedance (i.e. time/cost), yielding trip tables of person-trip demands. The relative proportions of trips made by alternative modes are factored into the model at the stage of *modal split*. At the final stage, *assignment/route choice*, modal trip tables are assigned to mode-specific networks. Generally, three different trip purposes: home-based work trips, home-based non-work trips, and non-home-based trips are defined in the model (McNally, 2000).

The dominance of the conventional FSM in producing aggregate forecasts as part of the transport planning process to date derives from its logical appeal, simplicity and tractability (Bates, 2000; Davidson et al., 2007). A fundamental conceptual problem with this approach, however, is its reliance on trips as the unit of analysis. As a trip-based approach, the FSM ignores the fact that travel is a derived demand; the motivation for the trips are, therefore, not explicitly modelled (Pinjari and Bhat, 2011; Malayath and Verma 2013, McNally, 2000). Given that different trip purposes are modelled separately, the scheduling and spatio-temporal interrelationships between all trips and activities comprising the individual's activity-travel pattern are not considered by the FSM (Dong et al. ,2006; McNally, 2000). Aggregate zonal analysis also implies that the effects of socio-demographic attributes of households and individuals as well as the behavioural complexities in travel captured in the FSM is limited (Martinez et al.,2007). This limits the ability of the approach to evaluate demand management policies and travel impacts of long-term socio-demographic shifts (Bhat and Koppelman, 1999; Pinjari and Bhat, 2011).

ii. Activity-based travel demand models

The activity-based approach (ABA) gained momentum around the 1990s with the promise of delivering a *behaviourally-oriented* alternative to the FSM. The conceptual underpinnings of this approach integrate aspects of the time-geography paradigm and human activity system analysis, as well as economic theory of consumer choice

The fundamental tenet of ABA is that travel is a derived demand; the need to travel is derived from people's desire to pursue various activities, which are interrelated (McNally and Rindt, 2007). The key areas of investigation in this approach, therefore, include the demand for activity participation, the spatio-temporal constraints within which activity-travel behaviour occurs, the complex interpersonal dynamics resulting from the interaction among household members and social networks, and activity scheduling and trip-chaining behaviour in time and space (Rasouli and Timmermans, 2014b, Bhat and Koppelman 1999; Yasim et al., 2016; Pinjari and Bhat, 2011).

Whereas a trip-based approach is satisfied with models that generate trips, ABA focuses on what generated the activities that in turn generated the trips through analysis of observed daily or multi-day patterns of behaviour (McNally, 2000, Dong et al. 2006; Lin et al., 2009). Contrary to the FSM, few activity-based models include route choice; activity-based models generate time-dependent O-D matrices, and if predictions of traffic flows are needed, these matrices serve as input to conventional route assignment algorithms (Rasouli and Timmermans, 2014a). The data requirements, model outputs, and fundamental principles of modelling travel demand using the FMS and/or ABA are not entirely different (Recker, 2001). However, the distinguishing feature of ABA relates to the "integrity, allowance for complex dependencies, higher resolution and time as a coherent framework" (Rasouli and Timmermans 2014b, p34).

All activity-based models adopt disaggregate modelling approaches. Most activity-based travel demand models including CARLA (Clarke, 1986), STARCHILD (Recker, et al., 1986), SCHEDULER (Gärling et al. 1989), TASHA (Miller and Roorda, 2003), AMOS (Pendyala et al., 2005) and ALBATROSS (Arentze et al., 2000; Arentze and Timmermans, 2004) are hybrid micro-simulation systems that combine a rule-based computational process approach with recent paradigms of agent-based modelling (ABM) to mimic how individuals build and execute activity-travel schedules. Others such as CEMDAP (Bhat et al., 2004), FAMOS (Pendyala et al., 2005) and CEMUS (Eluru et al. 2008) are utility-based econometric systems.

The activity-based paradigm has proven to pose serious impediment to the development of application models despite its conceptual clarity and purported unmatched potential for providing better understanding and prediction of travel behaviour (Recker, 2001). The approach is criticized for its lack of sound theoretical and rigorously structured methodological foundations (McNally and Rindt, 2007). Given that activity-travel decision processes have infinite feasible outcomes of many dimensions, modellers are presented with a fundamental combinatorial challenge (Ben-Akiva and Bowman, 1998; Rasouli and Timmermans 2014b) and several other problems related to the process of activity scheduling such as how utilities or priorities are assigned to activities and which heuristics and decision rules are used (Axhausen and Gärling, 1992**).**

2.6.3 Integrated land use and transport models

The key focus of LUTI Models is to integrate and predict four interdependent household choice sets namely; residential location, job location, vehicle ownership, and daily activity and travel patterns (Waddell, 2001). These models thus capture a range of spatial and non-spatial variables regarding geographical features and activity location, demographic and economic characteristics, transportation, and characteristics of the mobility demand (Eboli et al, 2012).

Over the past six decades, several operation LUTI models have been developed. These types of models are operational because they are full-fledged state-of-the-art models that have found practical applications in urban development and transport policy decision-making. Typically, an operational LUTI model has three main sub-model components, namely land use, socio-demographic, and transportation. These sub-models are either fully integrated or loosely coupled with each other to provide input-output linkages during model execution.

The land-use sub-model often contains important information on the urban land market, including residential and employment space ratio, land values, dwelling and occupancy types, land-use mix, housing vacancy, demolition and redevelopment (see figure 2.1). Most of the existing models (e.g., IMREL, KIM, MEPLAN, TRESIS, METROSIM, MUSSA, PECAS, RURBAN, TLUMIP, TRANUS, DELTA and URBANSIM) have detailed urban land and housing market sub-models.

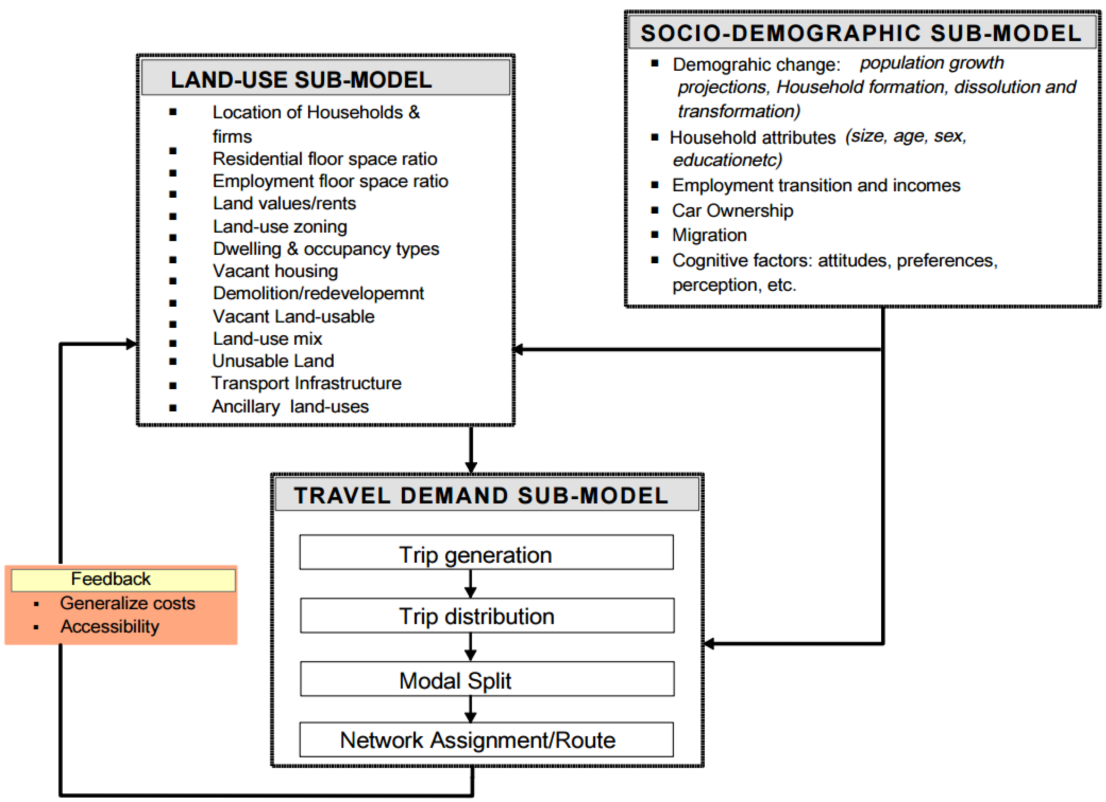


Figure 2.1: Structure of an operational LUTI model. Source: Acheampong and Silva, 2015, page 22

The socio-demographic sub-model contains important socioeconomic variables that mediate households' location choice and travel behaviour. Different model platforms have varying levels of detail they capture in terms of socio-demographic factors and processes. DELTA-START

(Simmonds and Still, 1999; Simmonds 2001) and UrbanSim (Waddell 2000), for example, have detailed demographic transition sub-models that capture the dynamics of household formation, dissolutions, and transformations as well as an employment transition model that simulates the creation and removal of jobs. At the household level, the demographic sub-models of most LUTI modelling frameworks often divide households into segments of similar socioeconomic groups. LILT (Mackett 1983, 1990, 1991), MUSSA–ESTRAUS (Martinez 1992, 1996) and RAMBLAS (Veldhuisen, Timmermans, and Kapoen, 2000) are based on 3, 13, and 24 different population segments respectively. Some operational models—DELTA-START and IRPUD (Wegener 1982, 1996, 2004) capture migration processes as part of their socio-demographic sub-models.

The transportation sub-model of most of the existing operational LUTI models, adopt the four-step approach. As shown in Figure 2.1, the land-use sub-model is dynamically coupled with the transportation sub-model containing a network assignment component. The extent and capacity of networks in the transportation sub-models for most LUTI models is held fixed or treated as a policy variable and, therefore, does not allow for evolutionary dynamics in transport networks (Iacono, Levinson, and El-Geneidy 2008). Generalized transport costs, manifested by congested networks, travel times, and distance are fed into the calculation of accessibility indexes, which in turn provide a dynamic feedback input into the land-use system.

The development of operational LUTI models has undergone waves of modelling techniques. Table 2.2 shows a classification of exiting operational frameworks according to modelling techniques; each column reflects the dominant theoretical and methodological persuasion of the model developers.

Early LUTI models were aggregate spatial interaction-based, drawing on the gravity analogy with entropy maximization as the underlying theory. In nearly all spatial interaction-based models, space is treated as discrete systems of aggregate zones; the zone systems afford the advantage of linking models with available data more easily and developing more mathematically tractable models (Pagliara and Wilson 2010). The need to capture complex individual behavioural dynamics and to overcome the weak assumptions and misspecification errors inherent in aggregate spatial interaction models have culminated in the adoption of disaggregate utility-based and micro-simulation methods in LUTI modelling. The metropolitan activity relocation simulator (MARS) adopts a somewhat different modelling approach. The model uses a systems dynamics approach in which a set of qualitative and quantitative tools are used to describe and analyse the dynamic feedback relationships between the land-use and transport systems and the underlying behaviour (Pfaffenbichler 2011).

Table 2.2: Operational LUTI models and modelling techniques

Aggregate Spatial Interaction-based Models	**disaggregate Utility-based Models**	**Micro-Simulation Models**	**Other**
ITLUP: DRAM, EMPAL, METROPILUS *(Putman, 1991, 1998)*	BASS / CUF Model *(Landis 1994; Landis and Zhang, 1998)*	ABSOLUTE *(Arentze and Timmermans, 2000, 2004)*	MARS *(Pfaffenbichler 2011, Mayerthaler et al. 2009)*— systems dynamics-based
KIM *(Kim 1989, Rho and Kim 1989)*	CATLAS, METROSIM (*Anas, 1983, 1984, 1994)*	ILUTE *(Miller and Savini, 1998; Miller et al., 2011)*	
Leeds Integrated Land-Use model (*Mackett 1983, 1990,1991)*	DELTA-START (*Simmonds and Still, 1999; Simmonds, 2001)*	ILUMASS *(Moeckel, et al., 2002)*	
Lowry-Garin model *(Lowry 1964)*	IMREL *(Anderstig and Mattsson, 1991, 1998)*	PECAS *(Hunt et al, 2008)*	
MEPLAN *(Echenique, 1969, Echenique et al,.1990)*	IRPUD *(Wegener, 1982,* 1996, 2004*)*	RAMBLAS *(Veldhuisen et al., 2000)*	
STASA *(Haag, 1990)*	MUSSA -ESTRAUS *(Martinez, 1992 1996)*	SIMPOP *(Bura et al. 1996, Sanders et al. 1997)*	
The Projective Land Use Model *(Goldner et al., 1972)*	RURBAN *(Miyamoto and Udomsr, 1996; Miyamoto et al 2007)*	TRESIS (Hensher and Ton, 2002)	
Time Oriented Metropolitan Model *(Crecine1964)*	Uplan *(Johnston et al., 2003)*	UrbanSim *(Waddell, 2000, 2002, Waddell et al. 2003)*	
TRANUS *(de la Barra, 1989, Donnelly and Upton, 1998)*			

Besides modelling approaches, the geography of application of the existing models is worth discussing. That is the spatial contexts in which models have originated or which models have been calibrated with data. Out of the 28 models reviewed, nine have originated from the United States (i.e., BASS/CUF, CATLAS, METROSIM, UrbanSim, Uplan, Lowry-Garin model, TOMM, Irvine simulation models, and TLUMIP). Three of the models have been applied in the Asian context: LILT and RURBAN in Japan, and MARS in Chiang Mai, Hanoi, and Ubon Ratchathani.

Moreover, three of the models (LILT, MEPLAN, and DELTA-START) have come from the United Kingdom. IRPUD, MEPLAN, and ILUMASS have been applied in the Dortmund region in Germany, while RAMBLAS and TRANUS have been applied in the Eindhoven region in the Netherlands and Curacao, La Victoria and Caracas in Venezuela, respectively. TRESIS has been used to investigate strategic-level policy initiatives for Sydney, Melbourne, Adelaide, Brisbane, Perth, and Canberra in Australia. Few of the existing models (i.e., LILT, ITLUP, MEPLAN, MARS and URBANSIM) have had large-scale international applications. ITLUP, a computer software for forecasting metropolitan spatial patterns of residential location and transportation, for example, has been calibrated for over 40 regions across the world. A thorough review of the literature on LUTI models revealed that not one of the existing LUTI models has either been developed in or calibrated with data from any African city.

2.7 Discussion of research directions and gaps in the literature

The preceding sections have reviewed the dominant theories and conceptual proportions, examined the empirical evidence, evaluated research approaches and methods and discussed how these have culminated in the development of various models of the relationship between urban land use and travel behaviour. This section provides a summary discussion of the key issues emerging from the literature, highlighting current research direction, methodological advances and outlining new areas for further research.

2.7.1 A shift towards disaggregate modelling approaches

One of the key issues emerging from the review of the literature is methodological. The literature review has shown that aggregate urban-economics and entropy-based gravity approaches are no longer adequate as standard frameworks for understanding the coevolution of the urban spatial structure and travel patterns. This is because models drawing on these theoretical and methodological approaches have several setbacks inherent in their aggregate and deterministic analytical approach, simplifying assumptions including monocentricity, spatial homogeneity, spatial equilibrium, unbounded rationality, and their inability to capture the richness of urban and regional geography as well as agent heterogeneity.

Instead, the current literature points to a shift from aggregate probabilistic modelling approaches towards stochastic disaggregate theory-driven modelling methodologies (Waddell, 2011; Chang, 2006; Iacono, et al., 2008; Pinjari and Bhat, 2011; Rasouli and Timmermans, 2014a). Three different modelling approaches namely, utility-based econometric approach, micro-simulation and ABM have emerged from the disaggregate modelling tradition. Among the disaggregate modelling approaches, the literature points to micro-simulation and ABM as novel simulation paradigm, grounded in complexity and systems theory, which make it possible to build spatially explicit and dynamic computational models of autonomous decision-making agents that interact with each other and their environment using a bottom-up approach (Railsback and Grimm, 2011; Batty, 2016). Thus, unlike their aggregate counterparts, these disaggregate approaches provide a framework to represent the complexities in individual choice behaviour as well as the attributes of their environment that shape their choices.

The emergence and dominance micro-simulation and ABM has been facilitated by technological advances which has made it possible to abstract and simulate complex systems such as cities in computer environments which hitherto was not possible. As revealed in the literature review, new generations of urban models continue to demonstrate the utility and capabilities of the disaggregate modelling tradition in general and ABM. Most importantly, the current generation of urban models adopting the ABM approach can put human decision-making behaviour at the centre of models and to relax the unrealistic assumptions of aggregate models by integrating various concepts and principles of urban location theory, utility theory, bounded rationality and heuristic decision-making into a composite framework to simulate human behaviour (Filatova et al., 2011; Huang et al., 2014; Batty, 2009).

Therefore, the adoption of disaggregate modelling such as micro-simulation and ABM approach to develop an urban location and mobility patterns model would constitute one of the major ways in which this thesis could make significant methodological contribution to the field of urban land use and transport modelling.

2.7.2. Understanding and modelling housing-job location choice decisions

The long-term choice behaviour of residential and job locations by individuals and households directly impact urban spatial structure and defines the set of activity-travel environment attributes available to a household or individual (Yang and Ferreira, 2008; Pinjari and Bhat, 2011). The ability to model residential mobility decisions has a great potential for improving long-range travel demand travel demand forecasting (Habib et al., 2011).

Effort to understand and model residential and employment location choice behaviour of individuals raises both empirical and methodological implications. Empirically, research drawing on access-space-trade-off principle derived from classical urban economics theory assume that workplace choice is predetermined or exogenous to residential location choice (Waddell et al., 2007). As the literature review has demonstrated, the exogenous work-place principle based on the assumption of conditional choice process for the entire population, is held in both the new generation of disaggregate urban location choice models adopting micro-simulation and ABM approach and some operational LUTI modelling frameworks (e.g., DRAM/EMPAL, CATLAS METROSIM, TRANUS, MEPLAN, and UrbanSim).

However, there is the possibility of a reverse choice process in which people select their places of residence before selecting their place of work (Waddell et al, 2007). Indeed, more recent empirical work (e.g., Boschmann, 2011; Habib et al., 2011; Pinjari and Bhat, 2011, Tilahun and Levinson 2013, Waddell et al. 2007, Yang, et al., 2013) has established that initial residential and job location choices as well as subsequent housing and job mobility decisions are jointly determined. A theory which does not depend on sequential ordering is useful because it allows for a better understanding of the interaction between the activities of individuals on the labour and housing markets (van Ommeren et al., 2000).

Thus, it is important for new research to observe heterogeneity in the population with respect to the different decision processes they use when deciding on where to live and where to work. The interdependence between residential and employment location choice decisions need to be investigated empirically in specific contexts to underpin the underlying assumption of modelling location choice in these contexts. As Lee and Waddell (2010) note, new empirical research from different contexts need also to move beyond the assumption of single worker households to understanding the residential-job location decision dynamics in multiple worker households. Further research is also needed in different contexts to better understand the effects of life-course events and changes in individual and household circumstances on job-housing location choice, the influence of households' most recent residence on evaluating future location choice as well as the job-housing location choice interdependence among multiple worker households (Lee and Waddell, 2010, Waddell et al., 2007).

Adopting a joint approach, however, presents the challenge of multi-dimensionality—a difficult analytical problem of modelling interdependence due to the many possible choice sets (Waddell et al., 2007). To handle the challenges of multi-dimensionality of choice alternatives a sequential ordering approach in which individuals search either for jobs given their residence, or individuals search for a new residence given the work place location may be adopted in standard in standard regression models (e.g. logistic regression models). The adoption of disaggregate micro-simulation or ABM could also help to relax the exogenous work-place assumption held in standard regression models and overcome the problem of multi-dimensionality by specifying residential-job location choice behaviour as interdependent without imposing a conditional structure on the decision-making process.

2.7.3. Integrating property market dynamics in disaggregate agent-based location choice models

Incorporating micro-level property market dynamics constitutes a critical element within the context of urban location modelling. This is because housing market processes determine how populations defined by various segments are distributed in space as well as the emergent activity and travel patterns (Ettema, 2011; Pagliara and Wilson, 2010). The ability of a model to capture market demand–supply interactions and to determine market prices endogenously is viewed as an issue of fundamental importance in assessing a given model's capabilities (Hunt et al., 2005).

The extent to which residential location models can realistically represent housing market transactions depends on several considerations. Firstly, agent heterogeneity, determined by the number of different segments of the population represented in the model, allows one to capture detailed socio-economic profiles as the basis of modelling complex decision-making behaviour.

Secondly, the types and attributes of choice alternatives that constitute the object of household agents' choice decisions, determines the model's ability to capture in rich detail, the attributes of the urban built environment. Thirdly, the ability to model the property development process although complex and computationally demanding, enhances model realism. Finally, improved realism is achieved in urban location choice models that can explicitly represent property market bidding processes (i.e. market interactions and transactions) and the resultant price setting outcomes that shape the interplay between demand and supply.

As demonstrated through the evaluation of urban location choice models, some attempts have been made to incorporate micro-level bilateral transactions and price formation to improve realism in new generation of disaggregate urban location choice models. Notwithstanding, there are several issues that require addressing in these models. The first improvement required in these models is the representation of spatial goods as discrete alternatives of households' choice. The review of the models found that, most them consider unique land parcels as the only discrete alternatives demanded by households. Although land matters in location decisions, it would be more realistic to include dwelling units differentiated by intrinsic attributes as these, in most cases, constitute the discrete alternatives of households' choice behaviour. Secondly, some models that include dwellings, do not simulate the full range of market tenure (owner-occupier or renting) and non-market tenure (living rent-free) options that might be available in real-world property markets (e.g. Ettema, 2011; Magliocca et al., 2011). These models tend to focus exclusively on market behaviour of potential buyers and owner-occupiers for that matter. However, not all households are buyers or would become owner occupiers. It is therefore essential for future model development to differentiate between different tenancies such as would-be-owner occupiers and renters and rent-free arrangements.

An equally important consideration with respect to providing realistic representations of property markets in location choice models relates to the way the supply side of the market is represented. A realistic representation of the supply side of the market will require simulating the development process by including the behaviour of different types of developers and how they interact with wider urban planning and development policies. Some of the more sophisticated operational LUTI models such as Urbansim include developers as agents in the urban development process. Other ABMs of land use (e.g. Murray-Rust et at., 2013; Hosseinali et al., 2013; Magliocca et al., 2011) also incorporate developers as actors in land conversion and development process. However, the supply side of the land and housing market is only passively represented in these models. For example, the constraints of urban development policy and zoning systems on land supply are not explicitly captured in these models. Moreover, the development permitting system which determines whether a proposed land use will take place or not is not included in such models.

It is worth mentioning that the role of private real-estate developers and the level of detail in which the activities of developers are modelled will depend on the purpose of the model. Also, private real-estate developers as actors in the property market may be important in certain contexts but not in all cases. For example, as will be discussed later in this thesis, in some contexts, including the case study are represented in this research, land development take place through households and individuals acquiring land and developing their own housing on incremental basis. In such cases, the way the supply side of the property market and hence the role of property developers are modelled will necessarily be different from when organized private real-estate developers are the main channels of land development and housing supply.

These considerations of model realism almost always increase model complexity, computation and tractability. However, detailed representation of property market dynamics including attributes of dwellings conditions of their environment at the meso- and macro- scales as well as bilateral transactions and competitive bidding endogenously determining prices, is crucial to developing realistic and robust models. As more property market elements are incorporated, the implications of a much broader range of policies, especially economic policies can be tested in the model, and this would potentially improve the capabilities of models as decision-support systems (Huang et al., 2014).

2.7.4 Explicit simulation of job location choice in disaggregate models of residential location to generate patterns of spatial interaction

The relationship between the work-place and the home, and the patterns of interaction between them is at the heart of LUTI models. Indeed, advanced operational LUTI models in addition to modelling location choice, can generate the associated mobility patterns of emergent spatial distribution of activities. New generation of ABM residential location choice models, however, do not yet have the capability of generating patterns of spatial interaction. As indicated by the literature review, this limitation is the result of the fact that none of these models of residential location choice explicitly simulates job location choice. Thus, to generate patterns of spatial interaction, these models must be extended beyond assuming the existence of an exogenously determined CBD used as reference point to determine cost parameters and proximity indexes as important factors influencing location choice. Instead, job location choice should be explicitly simulated as part of decision-making process. Tilahun and Levison (2013) have developed ABODE—Agent-Based Model of Origin and Destination Estimation in which they explicitly simulate job search and marching processes because of the interaction between firms and individuals. In ABODE, worker agents look for employment opportunities represented as firms distributed in the city, weigh offers and decide positions based on wages offered and their skills. The model however, takes the residential location of workers and the locations of employers as exogenous.

Given the advances in modelling micro-level residential location decisions on the one hand and the emerging innovation in explicitly modelling job search behaviour in models such as ABODE on the other hand, it is possible to develop new models that explicitly capture residential and job location choice decisions. Once the relationship between the work-place and the home locations are established in these models, it would then be possible to derive measures of spatial interaction such as origin and destinations pairs, trips distances, commuting time, mode use and accessibility indexes at micro and macro-levels in these models.

2.7.5 Expanding the geography of simulation model development and applications

The spatial and socio-economic contexts in which models have originated or which models have been calibrated with data is another key issue that emerged from the literature review. Firstly, the evaluation of the various empirical studies in the field of location choice and mobility characteristics revealed that considerable amount of fundamental research has long existed and continue to accumulate in cities within European and North American contexts. In recent years, cities in Asia, particularly China, have also received attention in land use and transportation model development. Consequently, the development and application of simulation models, as evidenced by the review of current state-of-the-art operational LUTI models have remained confined to these contexts. Neglected in both the literature of empirical evidence and model development are cities within the context of the Global South, especially in African cities which are experiencing rapid urbanization trends. In furtherance of the imperative to understand urban location and travel phenomenon in different contexts, new research that ventures into previously unexplored contexts would make significant contribution in expanding the current scope of research. Ultimately, the resulting research efforts would provide robust decision-support systems that would shape and inform complex urban development policy decision in these areas.

2.8 Formulation of research objectives and questions

The forgoing sections have provided an overview of previous research and highlighted gaps in the literature in relation to very fundamental empirical enquiries, methodological innovation as well as the realism and output capabilities of new generation of agent-based urban location models. Based on this discussion, the main objectives of this research and the accompanying research questions are outlined as follows:

2.8.1 Empirical research objective and questions

The first objective of this thesis is to:

- Examine empirically, the location choice behaviour of households and individuals with respect to their residential and job locations, and the mobility patterns associated with the observed home-work location combinations.

The set of fundamental empirical questions that would be pursued to address the above stated research objective are;

i. What are the factors and underlying processes of residential location choice behaviour of heterogeneous households in urban areas?
ii. What are the factors and underlying processes of job location choice behaviour of individual working members of the heterogeneous urban households?
iii. What are the interdependence between the residential location choice and job location choice of the households? and;
iv. What are the mobility patterns associated with the emergent residential-job location combinations?

2.8.2 Simulation model development objective and questions

The second objective of this thesis is to:

- Develop an integrated geospatial and agent-based model to simulate how the residential and job location choice behaviour of heterogeneous households and individuals co-emerge with mobility patterns in a metropolitan area.

The development of the model indicated in the above research objective was driven by the following specific research questions:

i. How do the socio-demographic characteristics and preferences of heterogonous households and individuals interact with existing urban structural conditions to influence urban location choice behaviour?

ii. How do bilateral transactions, competitive behaviour and interactions among individual actors in the property market lead to the formation and evolution of property prices?

iii. What are the residential location patterns that emerge from the interaction between households 'and individuals' choice behaviour and existing urban structural conditions?

iv. What are the employment location patterns that emerge from the interaction between the attributes of individual working members of the households and prevailing job market conditions?

v. How do the emergent residential and job location combinations and individual-level attributes of agents interact to shape the home-work mobility patterns?

The two main objectives outlined above are interrelated. The model development proposed would require that a set of fundamental empirical questions are addressed to implement it. To this end and in line with the gaps identified in the literature in relation to the need for the accretion of new empirical insights from different contexts regarding the residential-job location choice interdependence, the empirical research objective and questions are first addressed. The resulting analytical outputs, then serve as critical data inputs into the model implementation, calibration, verification and validation.

2.9 Chapter summary

In this chapter, previous research on the urban land use and mobility nexus have been reviewed. The review showed that research in the field is multidisciplinary and hence draws on theories and concepts from disciplines such as urban economics, urban geography, planning, social-physics and the complexity sciences. These theories, the review showed, have inspired a wide range of empirical research and simulation model development over the past six decades.

The literature review also showed that two major modelling approaches namely; aggregate and disaggregate modelling have evolved over the years. From each of these approaches, specific techniques and methods have also emerged. Urban economics and entropy maximization theory have underpinned the development of aggregate gravity-based spatial interaction models of urban location and spatial interaction. The development of random-utility and discrete choice theory made it possible for the development of utility-based econometric models that capture complex choice behaviour at the level of the individual who exercise choice with respect to finite number of discrete alternatives. Inspired by systems and complexity theories, microscopic modelling approaches such as agent-based modelling, have also been adopted in simulation model development.

The discussion of current research directions and areas of further research was framed around five main thematic areas. The first thematic issue being methodological, pointed to a general shift from applying aggregate modelling techniques towards the adoption of more innovative disaggregate modelling approaches. The need for empirical research from different contexts examining the residential and job location choice behaviour of households and individuals, and the interdependence between these choice sets was also highlighted. Moreover, the need to represent real-world property market dynamics to improve model realism; relax the exogenous work-place assumption in existing location choice models by explicitly simulating job location choice; and to expand the scope and coverage of future research beyond the confines of cities in Europe and North American contexts, by studying previously less explored contexts, were emphasized.

Following from the review of previous research and discussion of the research trends and gaps in the literature, the objectives and questions of this research were formulated. Two main research objectives, which reflect both the empirical and simulation model development focus of this thesis were formulated. Under each broad objective, specific research questions were also outlined.

In the next chapter, the overall methodology to that will be adopted to address the empirical objectives of this thesis will be discussed.

CHAPTER THREE: RESEARCH METHODOLOGY FOR EMPIRICAL STUDIES—CASE STUDY SELECTION, DATA COLLECTION AND ANALYSIS METHODS

3.1 Introduction

In chapter two, an account of previous research relevant to this thesis was presented. In addition to providing an overview of the general direction of research, the literature review identified and discussed the research gaps that will be addressed in this thesis. Based on the current direction of research and the gaps identified, two key objectives were formulated. Reflected in the two objectives, in broad terms, were the empirical aspects of the current research which would examine the urban location choice and mobility nexus in a specific urban context as well as the development of an urban simulation model to simulate how urban location choice co-evolves with mobility patterns.

To reiterate, the main objective of the empirical aspect of the research as stated in chapter two was to:

- examine empirically, the location choice behaviour of households and individuals with respect to their residential and job locations, and the mobility patterns associated with the observed home-work location combinations.

In relation to the above research objective, four fundamental research questions were derived to be pursued in this thesis. The key empirical research questions identified were:

i. What are the factors and underlying processes of residential location choice behaviour of heterogeneous households in urban areas?
ii. What are the factors and underlying processes of job location choice behaviour of individual working members of the heterogeneous urban households?
iii. What are the interdependence between the residential location choice and job location choice of the households? and;
iv. What are the mobility patterns associated with the emergent residential-job location combinations?

The focus of this chapter therefore is to set out the overall methodology adopted to address the empirical research objective and the accompanying research questions outlined above. To this end, this chapter discusses the case study design adopted and introduces the geographical scope—the case study metropolis for the empirical research. Anchoring the empirical investigation to this context, methodological issues including the delineation of the units of observation and analysis of the empirical research, the translation of the research questions into specific variables, research instrument design, sampling techniques, data collection, data analysis themes and statistical analysis methods are discussed.

3.2 Chapter organization

The rest of this chapter is organized into eight sections. The first section introduces the case study design and outlines the criteria and considerations for the selection of a suitable case study area for the empirical research. Following from this, the selected case study area would be introduced, highlighting its suitability using the established selection criteria. A detailed description of the case study area is then provided focusing on its location, size and historical growth process. In the fourth section, the rationale as well as the approach adopted to divide the case study area into broad urban-zones are presented. A programme to obtain the relevant data is presented in the fifth section, focusing on methodological issues including identification of study variables, data types and sources, description of sampling techniques adopted, design of research instruments and the administration of questionnaires. The penultimate section outlines the analytical themes derived from the data and specifies the corresponding statistical analysis methods that will be adopted. The chapter concludes with a summary of the methodological issues addressed in this chapter and opens a window into the subsequent chapters of the thesis.

3.3 The case study approach and case study selection criteria

The case study approach allows for extensive and in-depth analysis and description of a phenomenon within a given physical, socio-cultural, economic and political context (Yin, 2013). The case study design is appropriate when the phenomenon under investigation has both descriptive and exploratory aspects. As discussed in Hancock and Algozzine (2015), this approach involves the identification of research questions of interest and the determination of the appropriate unit of observation and analysis to answer the questions. It follows a systematic procedure of data collection from different sources and analysis of the data from which important findings about the case study are reported.

The selection of a suitable case study area is one of the crucial tasks involved in using the case study approach. While the very objective and questions being addressed by the research could influence the selection of case studies, other considerations such as data availability, peculiarities of a context that makes it an interesting case study area and familiarity with the context are important considerations too.

In this research, the selection of a case study area in which the empirical questions would be addressed hinged on a combination of objective and subjective considerations. As highlighted in the review of previous research in Chapter two, empirical research and the simulation models that have accompanied this research over the past five decades have focused largely on cities within Western European and North American contexts. Although research coverage in recent years is expanding into cities in the Global South (e.g. Hosseinali et al.,2013; Augustijn-Beckers et al, 2011; Barros, 2012; Xie et al, 2007), evidence remain limited particularly from cities in Africa. Given that cities in this region are among some of the rapidly urbanizing in the world, exhibiting varying degrees of complexity in their structure and evolution over time, detailed empirical research in these contexts would contribute to the accretion of knowledge towards a better understanding and theorizing of cities globally.

Moreover, data availability and access is very important to the successful execution of any research. While data in suitable formats, required to answer the empirical questions is almost always a challenge for much research, the collection of new data from primary sources is also constrained by time, personnel and financial resources. In addition to addressing the empirical research questions which constitute the focus of the current chapter, the development of the urban simulation model, indicated as the second objective of this research, details of which is provided later in chapters six, seven and eight of this thesis, requires large amount of data to provide the solid empirical foundations for implementation, calibration and verification. In the face of limited data and resource constraints, the researcher's familiarity with and tacit knowledge of the case study context, could be a useful asset to the overall research. Indeed, the researcher could take advantage of existing networks and contacts with institutions and individuals from whom data could the provided to inform the research.

3.4 The Kumasi Metropolis in Ghana as case study area: Location, size and historical growth patterns

The Kumasi Metropolitan Area (KMA) in Ghana, West Africa was selected as the case study area for this research based on the selection considerations outlined in the previous section. Specifically, the selection of the KMA as a case study area is consistent with the quest by this research to contribute to knowledge accretion and expansion by exploring and gaining insights into the urban structure and mobility nexus from an African metropolis. Thus, using this case study area, the unique physical, economic and socio-cultural factors shaping urban location decisions and the associated mobility impacts are examined while the extent to which the observed evidence support or contrast with already known evidence from other contexts are also highlighted.

Moreover, being a metropolis located in the researcher's home country, the researcher's background knowledge of the local context and existing networks and contacts, proved very helpful in the overall research design and in obtaining the relevant data from both primary and secondary sources for this research. As will be highlighted later in this chapter, the level of aggregation of existing secondary source data was not suitable for this research. In view of this, fine scale data at the household and individual levels was required. The collection of primary data at the household level benefited from background knowledge and existing networks of the researcher having lived in the metropolis for over twenty years and previously undertaken research work there.

Having introduced and discussed the suitability of the Kumasi metropolis as the case study area for this research, the sections that follow describes the case study area, providing information on its location, size, and growth patterns as important background to the overall design of the empirical study.

3.4.1 Location, size and historical growth patterns of the Kumasi Metropolis

The Kumasi Metropolis is a medium sized city located in Ghana, West Africa. The location of the metropolis, within the national, regional and sub-regional contexts in Ghana is shown in figure 3.1. The KMA is located between latitude 6°35" N - 6.40°N and longitude 1°30" W - 1°35''W, with an elevation which ranges between 250 - 300 meters above sea level. The metropolis is located approximately 270km north of Ghana's capital city, Accra. At the regional level, the metropolis is in the Ashanti Region, one of the 10 administrative regions of Ghana and functions as the administrative capital of the region. At the sub-regional level, the metropolis is located within the Greater Kumasi Sub-region (GKSR). The GKSR is functional region covering a contiguous area of approximately 2850km^2 comprising the Kumasi metropolis, the sub-regional core and its seven immediate surrounding districts as shown in figure 3.1.

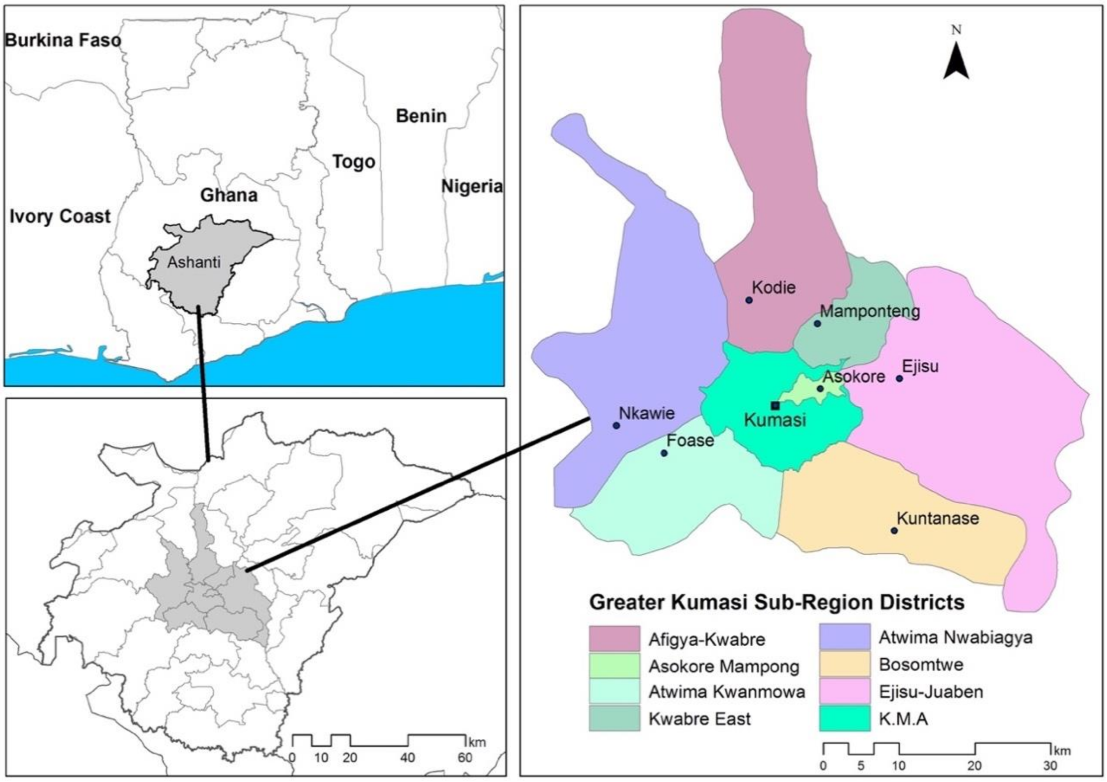


Figure 3.1: Location of the study area within the National, Regional and Sub-regional contexts

In terms of physical size, the KMA covers an estimated land area of 212km^2 representing nearly one percent and nine percent of the total land size of the Ashanti region and the GKSR. The Asokore Mampong Municipality, located to the east of the KMA (see figure 3.1) formed part of the administrative area of metropolis until it was later carved out as a separate administrative unit from the KMA in 2012.

What is known today as the Kumasi metropolis dates to the 17th century when the Asante Kingdom[2] was established by the *Asantehene* (King of the Asante State) at the cross-roads of

[2] The Asante Kingdom was a pre-colonial West African state/empire founded around 1670 by the Asante people—an ethnic subgroup of the Akan-speaking people in modern-day Ghana. The kingdom maintained a well-organized monarchical structure with power centred in the King, known as Asantehene, who ruled over small chiefdoms.

the Trans-Saharan trade routes. At the time, Kumasi, then a town of about 3,000 inhabitants became the political capital of the Kingdom (Amoako and Korboe, 2011).

Over the past one hundred years of its existence, the original town of Kumasi has expanded, merging into village settlements around its periphery. In 1995, under Ghana's decentralization program, the Kumasi Metropolitan Assembly, comprising the old Kumasi township and the surrounding communities, was established as an administrative unit by Legislative Instrument 1614. Today, the metropolis comprises the indigenous Kumasi township made up of the historical settlements of Adum-Kejetia, Asafo, Bantama and Manhyia, and over 90 distinct settlements. It is also one of only 6 metropolitan areas designated under Ghana's decentralization system comprising 216 local governments.

Historical population growth figures of the metropolis since 1984 are presented in table 3.1. From a base population of 487,504 in 1984, the metropolis experienced a four-fold increase in its population size to over two-million inhabitants in 2010, with current population density of about 5, 419 persons per square kilometre.

Table 3.1: Population size and growth rates in the KMA (1984-2010)

Districts	**Population Size**			**Annual Population Growth Rate (%)**	
	1984	**2000**	**2010**	**1984-2000**	**2000-2010**
KMA	487,504	1,170,270	2,035,064	5.63	5.69
GKSR	734,022	1,758,741	2,764,091	5.13	4.62
Ashanti Region	2,090,100	3,612,950	4,780,380	3.48	2.84
Ghana	12,296,081	18,912,079	24,658,823	2.73	2.69

Source: Based on 1984, 2000 and 2010 Population and Housing Census, Ghana Statistical Service

Growing at an annual rate of 5.69% since 2000, KMA is the fastest growing metropolis in Ghana. The average rate of population growth is higher than that of the national average (2.69%), the Ashanti Region (2.84%) and the GKSR (4.62%). Moreover, despite being the second largest metropolis, the KMA is growing faster than Accra, Ghana's capital city which has an average population growth rate of 4.2%.

The rapid population growth in the metropolis is reflected in the rate of built-up land expansion in the metropolis. Historical urban expansion patterns since 1986 are presented in Table 3.2. From an initial built-up land of 51 km^2 in 1986, the size of KMA's built-up land increased by three-folds to 177.501 km^2 in 2014 at an average annual expansion rate of 4.5%. Urban expansion in the metropolis has been characterized in recent years by sprawling development spreading from its historical core areas to the peripheral settlements within its boundaries as well as the neighbouring districts (Acheampong et al., 2016; Cobbina and Amoako, 2010; Amoateng et al., 2013). The rapid population growth and the attendant increase in built-up land, have implications for metropolitan spatial structure and the patterns of spatial interaction.

Table 3.2: Amount and rate of built-up land change in the KMA (1986-2014)

District/ Sub Region	**Total Area (km^2)**	**Built-up Land (km^2)**			**Annual Urban Expansion Rate (%)**		
		1986	2001	2014	1986 - 2001	2001 - 2014	1986 - 2014
KMA	212.093	51.756	108.007	177.501	5.027	3.895	4.500
GKSR	2849.933	87.970	176.499	400.516	4.752	6.506	5.563

Source: Acheampong et al. (2016, page 10)

For administrative purposes, the KMA is divided into nine sub-metropolitan units. Figure 3.2 shows the sub-metropolitan units with their respective physical whilst Table 3.3 provides a summary of the distribution of the metropolitan population and family size among the sub-metropolitan units.

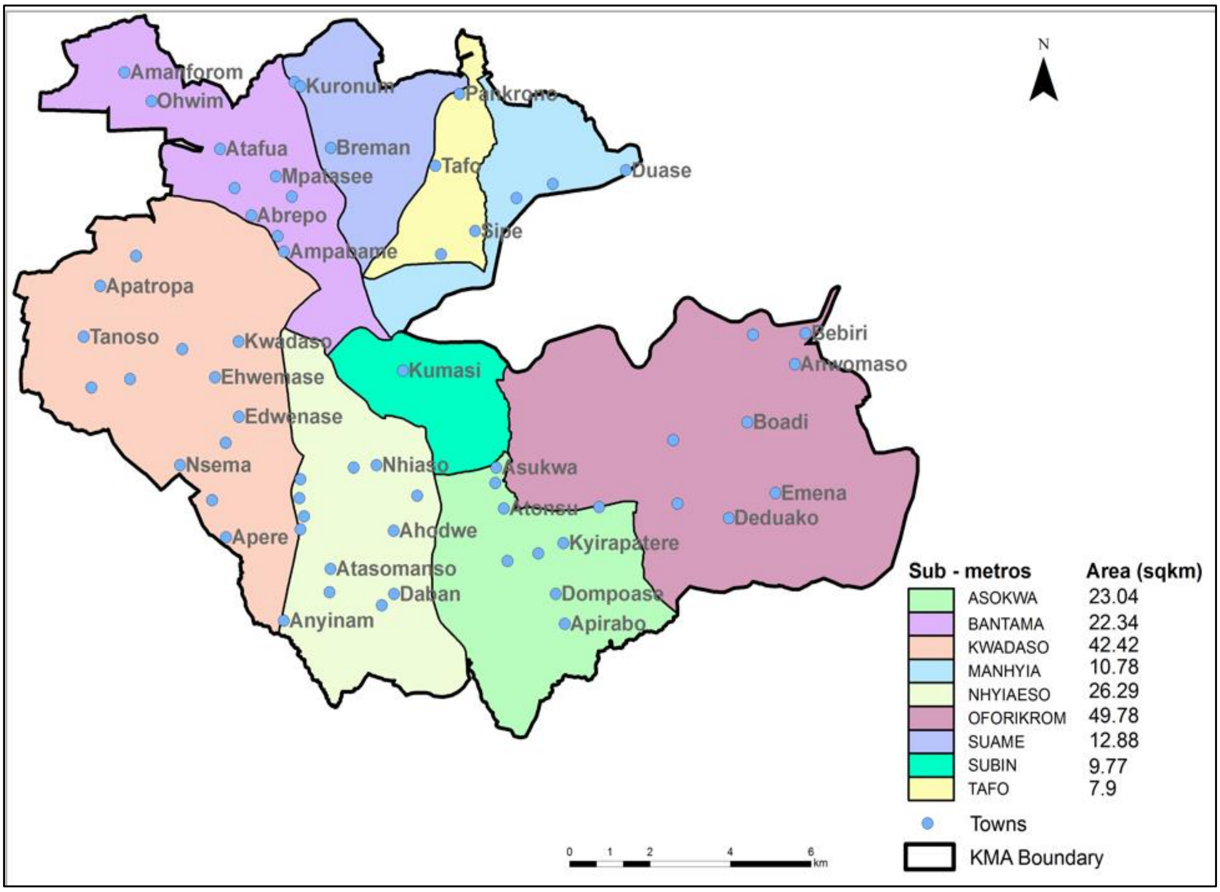


Figure 3.2: Sub-metropolitan divisions in the Kumasi Metropolis

Table 3.3: Population Distribution at Sub-metro level[3]

Sub-metro	Population	Percentage	Family Size
Asokwa	140161	8	3.87
Bantama	260474	15	4.09
Kwadaso	251215	15	4.17
Manhyia	152225	9	3.63
Nhyieso	134488	8	3.88
Oforikrom	303016	18	4.13
Suame	161199	9	6.91
Subin	174004	10	3.62
Tafo	146024	8	2.64
Total	**1,722,806**	**100**	**4.1**

Source: Based on 2010 Population and Housing Census, Ghana Statistical Service

3.5 Defining broad urban-zones in the Kumasi Metropolis

As mentioned previously, the nine sub-metropolitan divisions in the case study area exist for administrative purposes. These administrative units, although providing the basis for population distribution were not found to be particularly useful on their own for this research. In view of this, three broad urban-zones of unique physical and socio-economic characteristics were designated for this research.

The broad urban-zones were defined as successive circular zones extending outward from the central locations of the metropolis using three main criteria. The first criterion identified the historical origins of growth and expansion in the metropolis to define the first urban-zone hereafter referred to as the "*historical-core*". This zone covers traditional neighbourhoods within the inner ring-road system.

Development density analysis and broad area differentiation, drawing on the peri-urbanization literature on the metropolis were adopted to define the other two urban-zones, hereafter referred to as the *"Inner-suburban"* and "*Outer-suburban zones*". Firstly, using existing GIS data on all buildings in the metropolis overlaid on a layer of square grids of 0.25km^2, a point density analysis was computed as shown in figure 3.3. Secondly, the peri-urban zone of the metropolis, which has previously been mapped informed the designation of the zones. The urban-periphery is conceptualized as a transition zone between fully urbanized land in cities, and areas in

[3] This excludes the population of Asewase sub-metro (312,258), the 10th sub-metro which now forms part of the newly created Asokore-Mampong Municipality. Together, this makes a total population of 2,035,064, provided by the 2010 census data.

predominantly agricultural use, characterized by mixed land uses and indeterminate inner and outer boundaries, (McGregor et al., 20011; Webster, 2002). Simon et al. (2004), estimates that the peri-urban zone of Kumasi, for example, stretches some 20 to 40km radius around the city's main built-up area. All towns within this zone have been mapped. Thus, in delineating the inner- suburban and outer-suburban zones, the geocoded location of all towns in the metropolis were overlaid on the density map. Using the names of the peri-urban settlements as the main criteria and the results of the density analysis as a further check the boundary of the outer-suburban zone was distinguished from the inner-suburban zone. The outer-suburban zone comprises peri-urban settlements within the metropolis where development density is generally low between 35 and 136 buildings per 0.25km$^{2.}$ The inner-suburban zone covers areas of relatively higher development density (i.e. 137-530 buildings per 0.25km^{2}.) that extend between the historical-core and outer-suburban zones as depicted in figure 3.4. It is worth noting that in the density map (i.e. Figure 3.3) cells with less than 35 buildings per 0.25km^{2} are mostly areas covered by institutional buildings, nature reserves and parks, and the University (i.e. KNUST) located in the eastern part of the metropolis within the outer-suburban zone".

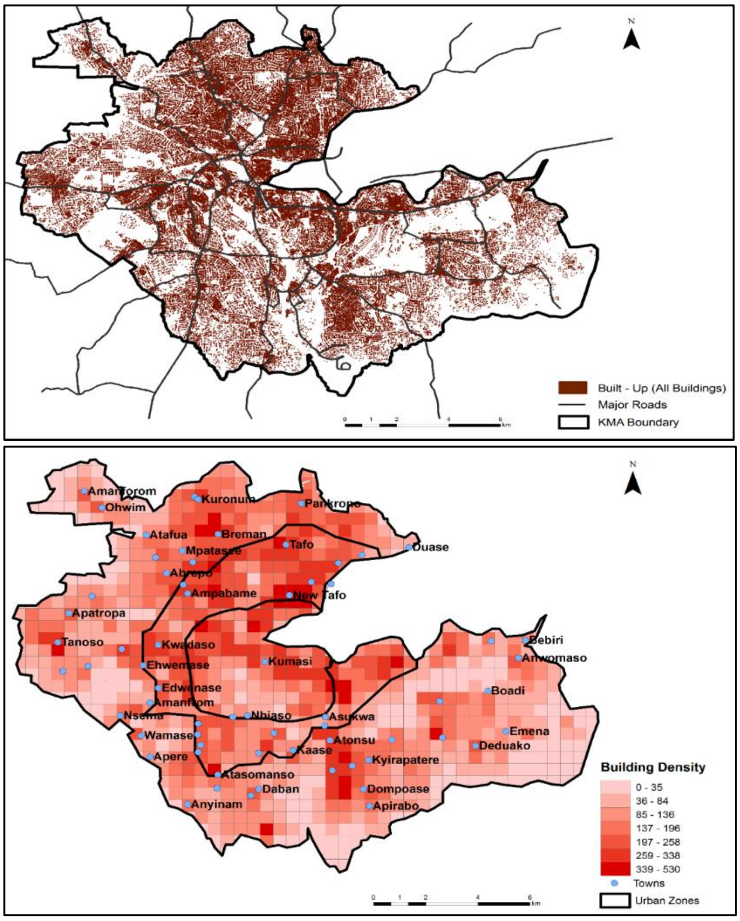


Figure 3.3: Maps showing location of buildings and building density

in the Kumasi metropolis: Based on data obtained from the TCPD, KMA

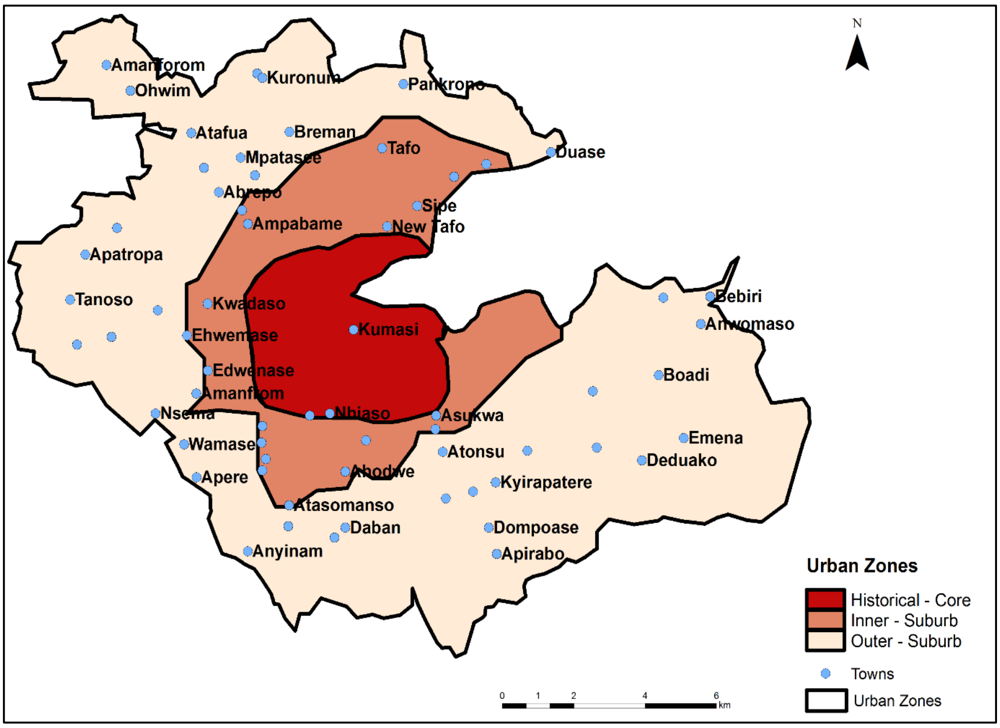


Figure 3.4: Map showing broad urban-zones in the Kumasi Metropolis
Source: Author's construct based on data obtained from the TCPD, KMA

Covering a total area of 22 km^2 (11% of the total land area of the metropolis) the historical-core zone comprises the centrally located traditional neighbourhoods of the metropolis. Also, located in this zone is the metropolitan Central Business District (CBD). The inner-suburban zone covers a total land area of 38.7 km^2 of contiguous high to medium density development immediately surrounding the historical-core. It comprises approximately 19% of the total land area of the metropolis. Finally, the outer-suburban zone covers a total land area of 145km^2 of medium to low density contiguous urban land immediately surrounding the inner-suburban zone.

As will be demonstrated later in this chapter, the three urban-zones designated here together with the existing eight sub-metropolitan administrative units provided the spatial units for the multi-level sampling approach adopted during data collection. Also, reference is made to these urban-zones throughout the remaining chapters of the thesis as distinct spatial units for analytical purposes.

3.6 Survey design and data collection

Having set the geographical context for this research in the previous sections, this section details the survey design and data collection activities carried out as part of this research. It addresses the unit analysis of the research and identifies the data types and variables, and the corresponding data source in line with the research questions.

3.6.1 Units of observation and analysis

The unit of analysis refers to the primary entity under study. As explained by Yin (2013), it is the 'who' or 'what' that is being studied. These entities could be individuals, households or organizations observed at different spatial scales in line with the data requirements of the research.

In accordance with the objective and research questions, two main entities namely, households and individual workers in the household constituted the unit of analysis for the empirical research. The research adopted the official definition of the household used by the Ghana Statistical Services as comprising a person or a group of persons, who live together in the same house or compound and share the same house-keeping arrangements. These entities are the focus of the research because residential location choice decisions are made by households while adult individuals within the household make job location choice decisions as well as daily travel decisions.

It is worth mentioning that, institutions and agencies with competencies in making various policies that shape population, employment, infrastructure development and physical organization of functional land uses in the metropolis were consulted for purposes of data collection. These were not units of analysis per se. Instead, they were sources of secondary information to inform the research. Indeed, as will be explained later, the information from these institutions were at very coarse resolutions and therefore only provided a starting point to gathering the fine-scale data directly from the households and individuals that constitute the primary unit of analysis of the empirical research.

Furthermore, the location choice decisions of the households and patterns of home-work mobility of individuals were observed at different spatial scales. For example, residential location choice was examined at the macro-scale where location defining attributes including the location of employment and amenities were evaluated by the households; at the meso-scale, where the focus was on the three broad urban zones—historical-core, inner-suburb and outer-suburb; and at the micro-scale where attributes of dwellings such as dwelling type, tenancy, size and rent are considered. Similarly, mobility patterns were observed at the level of individual workers based on the locality of home and employment locations and anchored to existing Traffic Analysis Zones (TAZs), comprising of aggregate urban-zones containing many activities to represent flows and patterns of spatial interaction.

3.6.2 Secondary data from institutions and agencies

The data collection exercise began with a survey of institutional sources in Ghana for secondary data that could inform this research. Specific datasets required for this research were first identified followed by establishing contacts with the relevant institutions to ascertain what aspects of the required data exist.

As outlined in table 3.4, datasets obtained from institutional sources covered demographic, socio-economic and physical conditions of the metropolis. Aggregate population data at the national, metropolitan, sub-metropolitan and urban settlements scale, was obtained from the Ghana Statistical services in the format of census reports. Data on urban land use and activity locations, and administrative boundaries and infrastructure were obtained mainly from the Town and Country Planning Department and Urban Roads Department.

Table 3.4: Summary of data obtained from institutional sources.

Datasets	Sources (Institutions, Agencies)	Format	Spatial scales
1. Historical population and housing data	Ghana Statistical Services	Reports: Ghana Population and Housing Census (1986, 2000, 2010)	Aggregated at national, metropolitan, sub-metropolitan and settlement scales.
2. Urban land use development trends	Town and Country Planning Department, Kumasi	Report: Comprehensive Urban Development Plan for Greater Kumasi, January 2012	Aggregate land use distribution data at the metropolitan scale.
3. Socio-economic development trends	Development Planning Unit, Kumasi	Report: Medium Term Development Plan, Kumasi Metropolitan Assembly, 2012	Aggregate information on population, employment and socio-economic conditions at metropolitan and sub-metropolitan scales
4. Historical land values data	Lands Commission, Kumasi, Ghana	Raw land transactions data from 2000, 2005, 2010 and 2015	Available at settlement level. Request was made for land value data for 30 proxy locations in the metropolis
5. Metropolitan Traffic Analysis zone system and travel data	Department of Urban Roads, Kumasi, Ghana	Report: Urban Transport Planning and Traffic Management Studies for Kumasi, August 2004	Analogue diagrams of TAZ system obtained to be digitized in GIS as well as origin destinations data in 2004.
6. Administrative boundaries data	Town and Country Planning Department, Kumasi	GIS shape file format	National, regional, sub-regional, metropolitan and sub-metropolitan scales
7. Housing location dataset	Town and Country Planning Department, Kumasi	GIS shape file format	Metropolitan, sub-metropolitan and settlement scales
8. Infrastructure and amenities location (road network, transport terminals, schools, markets)	Department of Urban Roads, Kumasi, Ghana Town and Country Planning Department, Kumasi	GIS shape file	GIS data of road hierarchy: arterials, distributors and access roads, transport terminals, schools at metropolitan, sub-metropolitan and settlement scales
9. Natural urban features (rivers, forests)	Town and Country Planning Department, Kumasi	GIS shape file	Metropolitan, at metropolitan, sub-metropolitan and settlement scales

Although data obtained from these institutions provided the relevant contextual and situational information about the case study area, there were several limitations with them. Except for the GIS data that was current, most of the information obtained from reports were out-of-data for this research. For example, the only available travel related information obtained from the Urban Roads Department came from a study that was conducted in 2004, more than 13 years ago. Also, the existing socio-economic data (i.e. population, housing and employment) were aggregated mainly at the national, metropolitan and sub-metropolitan scales and did not contain the relevant variables at the level of the household required for this study. Data on residential and job location choice, at the level of households and individuals, the main subject of this research was non-existent.

In view of the limitations of the data from the institutional sources, a survey to obtained fine scale primary data from households and individuals on their residential and job location choice as well as travel patterns was necessary. The section that follows describes the primary data collection activities.

3.6.3 Primary data collection: Identification of study variables based on research questions

As the basic unit of analysis for this research, households and individuals within the households were the source of primary data. To obtain the relevant primary data, the four research questions driving the empirical studies were translated into specific analytical themes from which the specific variables were identified. Table 3.5 provides an outline of the research questions, the corresponding analytical themes and the key variables required to address them. The primary data required, consisted of disaggregate quantitative and qualitative variables representing the demographic and socio-economic attributes of households, housing and land market characteristics, residential and job location preferences and as well as mobility characteristics.

Table 3.5: Linking research questions to analysis-themes and study variables

Research Questions	Analysis themes	Key Variables
1. What are the factors and underlying processes of residential location choice behaviour of heterogeneous households in urban areas?	i. Analyse of background socio-demographic characteristics of households	▪ Marital status, family size, age and gender of household members, number of working members, income, educational attainment levels.
	ii. Examine characteristics of the metropolis' housing and land market	▪ lot size, house types, house size (number of bedrooms) tenancy types, house rents/prices land values.
	iii. Identify the macro and meso scale residential factor preferences of households	▪ Households' attachment of importance to proximity to urban infrastructure, essential amenities, safety, family relations and social networks, neighbourhood characteristics in residential choice.
	iv. Examine the determinants of preferences for housing types	▪ Dwelling types, income-group, marital status, family size, educational attainment levels, urban-zones of residence.
	v. Examine the determinants of preferences for housing tenancy	▪ Housing tenancy, dwelling types, income-group, marital status, family size, educational attainment levels, urban-zone of residence.
	vi. Assess house rent-to-income ratio among households	▪ House rent bands, household incomes
2. What are the factors and underlying processes of job location choice behaviour of individual working members of the heterogeneous urban households?	i. Identify characteristics of the job market	▪ Employment industry of individuals, primary and secondary occupation types,
	ii. Examine the spatial distribution of jobs	▪ Locality of work-place location, home-based job location, non-home-based job locations, urban-zone of employment location, major employment zones of job location.
	iii. Identify the macro and meso scale job location factor preferences of individuals	▪ Individual's attachment of importance to proximity to urban infrastructure, essential amenities, and wage opportunities in job location choice
	iv. Analyse of the determinants of Job location	▪ Job locations, income-group, educational attainment and skill levels of workers, work Industry, urban-zone of residence.

Table 3.5 continued: Linking research questions to analysis themes and study-variables

Research Questions		Analysis themes		Key Variables
3. What are the interdependence between the residential location choice and job location choice of the households	i.	Examine the relationship between the residential and job location choice of households	▪	Residential and job location choice decision ordering of households, households' residential location change and reasons for change of residence, individuals' job location change, and reasons for job location change
4. What are the mobility patterns associated with the emergent residential-job location combinations?	i.	Analyse home-work trip production and attraction among traffic analysis zones	▪	Employment locations, home locations, Traffic analysis zoning system,
	ii.	Examine the determinants of private car ownership	▪	Vehicle ownership rate, income, educational attainment of household heads, marital status, family size and composition, distance of residence to CBD, distance between home and work, urban-zone of residence
	iii.	Examine the determinants work transport mode choice	▪	Income, available transport modes, considerations for transport mode use, urban-zone of residence, job location, distance between home and work
	iv.	Analyse travel times and costs associated with home- work commuting and their determinants	▪	Work trip frequency, mode choice, commuting times, out-of-pocket transport costs, incomes, home-work distance

3.6.4 Questionnaire design

The key study variables identified from the research questions in table 3.5 were translated into a structured questionnaire. The research questionnaire comprised six thematic sections, each containing a set of closed and open-ended questions used to capture the relevant data. A brief description of the thematic sections is provided below.

i. Background socio-demographic and socio-economic information

The first theme of the questionnaire contained questions used to obtain background information about household characteristics including marital status, family size and composition, levels of educational attainment of members and employment status and income. This data would provide the basis to differentiate households and individuals as heterogeneous entities and to determine how these attributes shape their location and mobility choices.

ii. Housing market characteristics and residential location preferences

The second theme of the questionnaire contained questions used to elicit data about households' locality and urban-zone of residence, dwelling types occupied, tenancy arrangements and rent levels paid by those in the rental sector. Questions aimed at gaining data on recent residential location change and the reasons for the change were also included in this section.

In addition, the residential location preferences of households were elicited under this theme. Questions were formulated to allow households to give a retrospective account of the factors they considered important in deciding their current residential and job locations. Based on a thorough review of the literature and the researcher's knowledge of the local context, 19 residential location factors were identified. These factors were used to constitute measurement items to which households would indicate the importance they attach to them when deciding their current places of residence. Based on knowledge of the local context that extended family relations and social networks influenced residential location choice, items that reflect these factors were included in the questionnaire. Response to each of the items was constructed on a 5-point Likert scale ranging from *"very important"* to "*not important at all*". A sample of the residential location preference items in the questionnaire is provided in figure 3.5. In addition to the Likert scale items, open ended questions were used to elicit additional information about the reasons for households' choice of their current residential locations as well as the reasons underpinning their residential location changes in the past.

A. Residential Location Choice

Preamble: On a scale of 1 to 5 please indicate the importance to you of each of the following factors when deciding your current place of residence, where 1 is very important and 5 is not important at all. *(if any do not apply, tick 3-Neutral or Not Applicable)*

Factors/Reasons	1 Very important	2 Important	3 Neutral/NA	4 Not-important	5 not-important at all
Prestige of the neighbourhood					
Proximity to my workplace					
Proximity to workplace of my partner/spouse					
Proximity to major road					

Figure 3.5: Sample of questionnaire items used to elicit residential location choice and preferences

iii. Job market characteristics and job location preferences

The third theme of the research questionnaire comprised questions used to obtain data on job market characteristic including the employment industry and occupational types of adult working members of the household, the number of jobs done by workers, the locality of employment location and job location type (i.e. home-based and non-home-based locations). Questions aimed at gaining information on recent job location change and the reasons for the change were also included in this section. Analysis of this alongside residential mobility data would allow understanding of the residential-job location choice interdependence

To understand the job location factors individual adult working members in the household considered important, 8 factors identified from the literature as influencing job location choice were formulated as questions. A 5-point Likert scale ranging from *"very important"* to *"not important at all"* accompanied each of the factors. A sample of the job location preference items in the questionnaire is provided in figure 3.6. In addition to the Likert scale items, open ended questions were used to elicit additional information about the reasons for individuals' choice of their current job locations as well as the reasons underpinning their job location changes in the past.

B. Job Location Choice					
Preamble: On a scale of 1 to 5, please indicate the importance to you of each of the following factors when deciding your current job location, where 1 is very important and 5 is not important at all. ***continue on supplementary sheet if working members exceed 2***					
Factors/Reasons	1	2	3	4	5
	Very important	Important	Neutral/NA	Not-important	not-important at all
Proximity to home/residence					
Opportunities for high paid jobs					
Proximity to work place of my spouse/partner					

Figure 3.6: Sample of questionnaire items used to elicit job location choice and preferences

iv. Mobility characteristics

Mobility characteristics, constituted the fourth theme in the questionnaire, which was linked to the fourth research question of the empirical study. Under this theme, questions were formulated to obtain data on home-work mobility patterns, transport mode use, travel frequency and travel time. The specific questions focused on work trips start and return times over a 5-day period and the reasons for travel time start times. Questions on transport mode for work purposes over a 5-day period were also included. Six items intended to capture attitudes and considerations for travel mode use namely; predictability, affordability, comfort, privacy, flexibility and travel time, with each accompanied by a 5-point Likert scale ranging from “very important” to “not important at all were formulated to be evaluated by individuals.

v. Household income and expenditure

The final theme of the questionnaire comprised questions used to elicit data on the income and expenditure of households. The main items here included monthly income of all adult working members of the household and spending on housing, transport and other non-housing and non-transport related expenditures. The affordability perceptions of households in relations to housing costs and transport costs were also derived based on questions included in this section of the questionnaire.

3.7 Conducting the household survey

Having delineated the research unit of analysis, identified the key variables and data sources, and designed a questionnaire, the next step of the research process was to conduct a household survey. The household survey involved important methodological considerations including sample size determination, recruitment and training of field assistants and the conduct of actual household interviews. Details of the activities carried out under each stage of the survey process is presented in the subsequent sections.

3.7.1 Sampling technique

A multi-level sampling approach, involving the determination and allocation of the representative sample of households to be interviewed at the different spatial scales was implemented. An overview of the sampling procedure is presented in figure 3.6. The process is broken down into two steps involving the determination of sample size from a sample frame of households and the distribution of the selected sample among the spatial units identified in the Kumasi metropolis.

Based on the population data obtained from the Ghana Statistical Services, a total of 43, 6691 households in the Kumasi metropolis, constituted the sample frame. The sample size of households was determined using the formula below:

$$\boldsymbol{n} = \frac{\boldsymbol{z}^{2} \times \boldsymbol{p}(\mathbf{1} - \boldsymbol{p})}{\in^{2}} \qquad \mathbf{3.1}$$

Where:

n = sample size

z^2 = z-score

p = population proportion

$\in$ = margin of error

A Confidence level of 99%, which is a measure of the certainty regarding how accurately the sample size reflects the population was assumed. A 99% confidence level has a corresponding z-score value of 2.58. The z-score assumes that the sampling distribution is normally distributed. A population proportion of 0.99 was assumed. Plucking these values into equation 3.1, the total minimum sample size of 665 was determined. The sample size calculation therefore yielded a total household sample of 665 to be interviewed in the Kumasi metropolis.

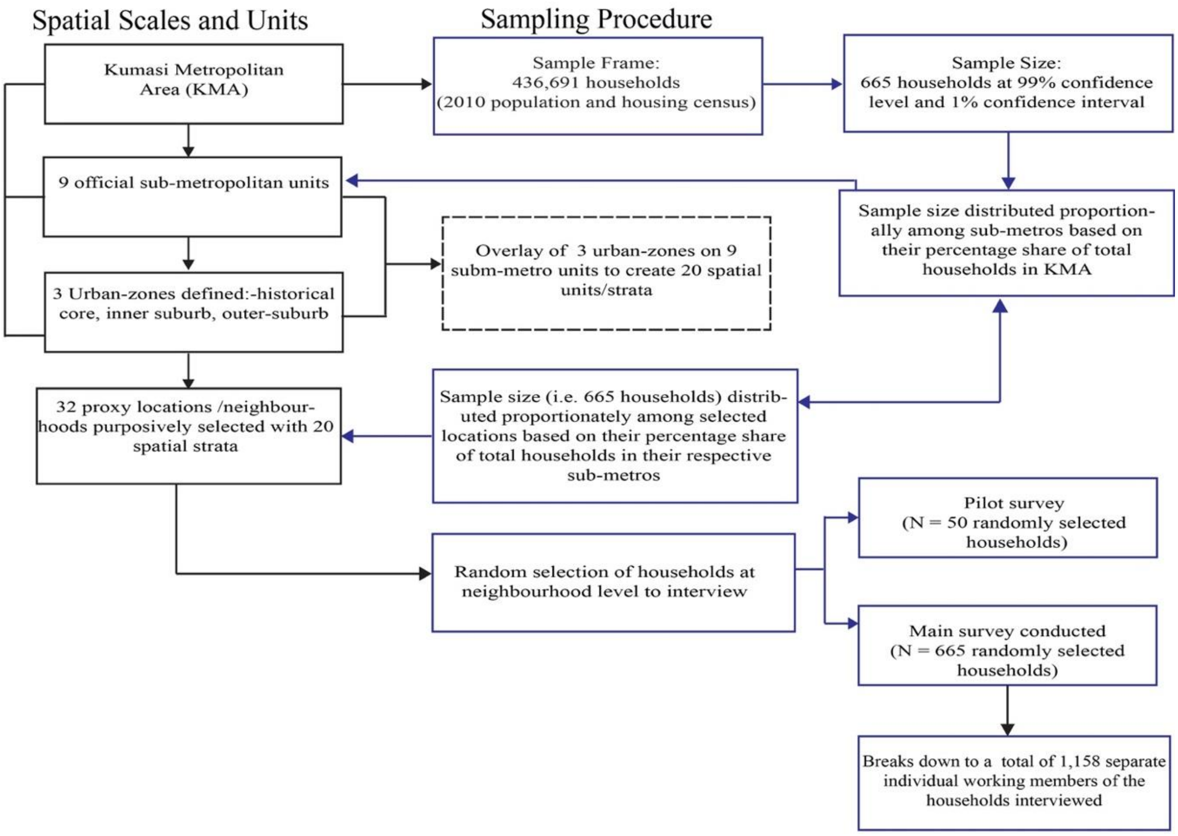


Figure 3.7: Multi-level sampling procedure

Next the total sample size was allocated proportionately among the spatial units at different scales. At the first spatial scale of the sample size distribution, each of the nine sub-metropolitan units in the Kumasi metropolis was allocated a share of the sample size as a proportion of their respective households to the total number of households in the metropolis. The sample distribution at the sub-metropolitan scale is summarized in table 3.6. For example, since Asokwa sub-metro has a total household of 36,183, representing eight percent of all households in the metropolis, it was allocated 55 (i.e. eight percent) of the total 665 sample size.

Table 3.6: Sample frame and sample size distributions at the sub-metropolitan scale

Sub-metros	**Same frame**		**Sample size distribution**	
	Total Households	Percent of total households	Sample	Percent of Sample
Asokwa	36183	8	55	8
Bantama	63722	15	96	15
Kwadaso	60233	14	92	14
Manhyia	41886	10	64	10
Nhyieso	34624	8	53	8
Oforikrom	73343	17	112	17
Suame	23318	5	36	5
Subin	48105	11	73	11
Tafo	55277	13	84	13
Total	**436691**	**100**	**665**	**100**

At the second stage of the multi-level sampling process, proxy settlements were selected from each of the sub-metropolitan units to which the initial sample size allocation was further distributed proportionately. To ensure fair representation at the settlement level, the three urban-zones designated were overlaid on the existing sub-metropolitan units. This resulted in 20 spatial units or strata being established. Next, all the settlements within the metropolis were superimposed on the 20 strata. From this list, a total of 32 proxy locations were identified as the household data collection points in the metropolis. These were the major settlements in terms of population size within the 20 spatial strata created.

Each of the selected settlements' share of households to that of the sub-metro in which it was located was computed. Using these percentages, the sample size allocated to each of the sub-metros at the first stage of the multi-level sampling process were proportionately allocated to the selected settlements under each of them. The distribution of the sample size at the settlement level within the sub-metropolitan units and broad urban-zones is depicted in figure 3.8. The number of questionnaires allocated to each settlement is indicated in brackets in figure 3.8. Overall, 179 households (i.e. 20% of the total sample size) were selected from 8 proxy locations within the historical-core area of the metropolis. A total of 285 households (44% of total sample size) from 11 proxy locations and 201 households (36% of total sample size) were selected from the inner-suburban and outer-suburban zones respectively.

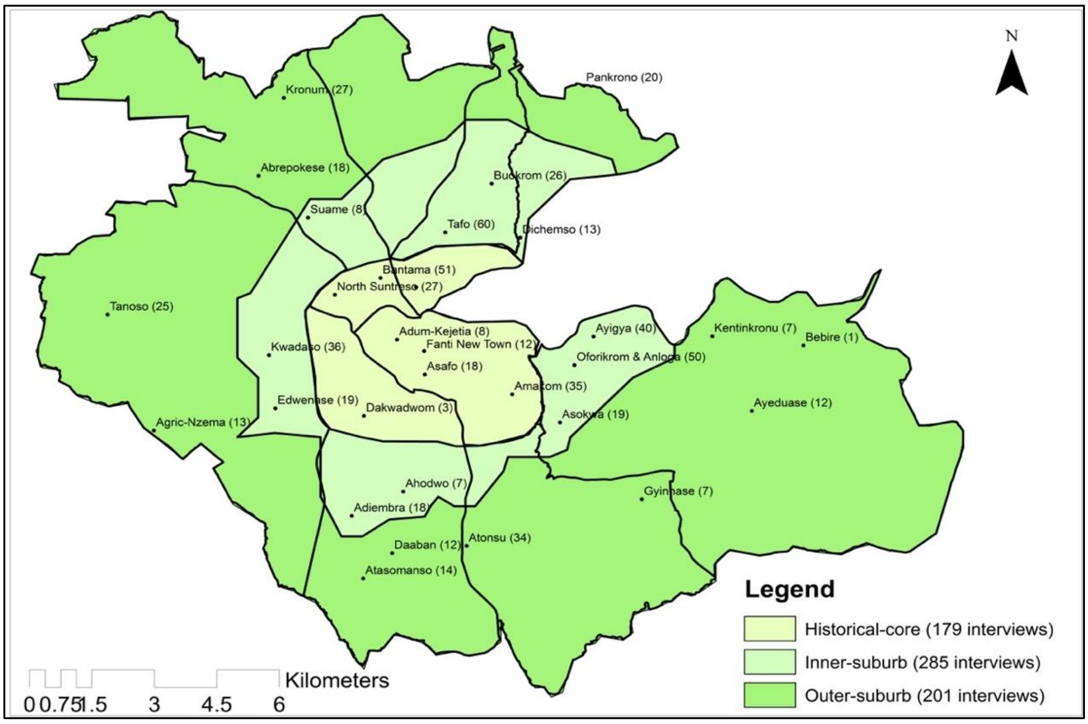


Figure 3.8: Distribution of sample size into proxy communities within sub-metros and designated urban-zones

3.7.2 Recruitment and training of field assistants

Given the scope of the fieldwork, which involved interviewing 665 households across the metropolis, field assistants were recruited and trained to assist in conducting the interviews. A total of 20 field assistants were recruited for the survey. These were all graduates who had completed the undergraduate degree program in Planning at the Kwame Nkrumah University of Science Technology—the researcher's previous university in Ghana. Their background training and experience with questionnaire administration and the fact that all of them were resident in the Kumasi metropolis made them suitable as field assistants for this research.

Four training sessions, over a 2-week period were organized for the research assistants as a group. The first group training session involved introducing the background to the research and the overall objectives of the research to the field assistants. In subsequent sessions, the field assistants were taken through each of the questionnaire items. The aim was to explain each question and the response it was intended to elicit to them. Also, since the survey instrument was constructed in English, each of the questions had to be translated to Twi—the local language spoken by residents of the metropolis. In the process, possible ambiguities in the meaning of sentences were clarified whilst a uniform translation of each questionnaire item from English to the local dialect was arrived at and noted by the field assistants. Copies of the questionnaires were given to the field assistants to study in detail for one week after which the last training session was held to address possible questions and to provide further clarification. All the field assistants were paid an agreed fee covering their services and transport costs.

3.7.3 Pilot interviews

A pilot survey was conducted with the aim to test the questionnaire and to afford the field assistants the opportunity to gain further understanding and familiarity with the instrument before the actual interviews. A total sample of 50 randomly households were interviewed in the pilot survey. The results of the pilot survey influenced the final questionnaire in two major ways: First, it became evident that the initial questionnaire was quite long. Consequently, it was shortened by reducing the number of items on it. Second, some of the questions were found not to be very clear to the survey respondents and were modified accordingly. The modified questionnaire designed was emailed to my supervisor for approval. The approved questionnaire was the used for the actual survey.

3.7.4 Conducting the household interviews

With a questionnaire designed, a representative sample size of households determined and allocated proportionally among 32 proxy locations across the metropolis, the next stage of the survey process involved selecting and interviewing the households at the settlement level. At the settlement level, the selected communities were further differentiated into the indigenous neighbourhoods and the surrounding, relatively new residential neighbourhoods. The allocated sample size for each settlement was then distributed equally between these two areas.

In selecting households to interview, residential blocks were identified using the existing street system in the neighbourhoods. From each residential block, a house was randomly selected until the required sample was reached. From the selected house, a household was identified and interviewed. Where the selected house was occupied by multiple households, only one household was interviewed there. Households who were at home as of the time their houses were visited and were willing to participate in the study were selected to be interviewed. To have a good representation of the population, the interviews were conducted during the day and evening times on both week-days and weekends. This way, there were better chances of reaching household members who were involved in home-based work and those involved in non-home-based work.

All interviews proceeded on a prior agreement between the interviewers and interviewees that the responses elicited were for research purposes only and that they would be anonymized. Within the selected household, interviews were conducted at two levels. Firstly, data on the socio-demographic attributes of the household as well as their residential location preferences was obtained from the household heads. At the second level, all adult working members of the household were identified and interviewed on their job location choice and travel characteristics. Thus, although the interviews started with a target sample of 665 households, at the end, 1,158 individuals from the targeted households were interviewed. All interviews involved face-to-face interaction sessions with the respondents during which the interviewer read the interview questions to them and manually recorded their responses on the questionnaire.

The reliability of the responses elicited from the households and individuals depends on several factors. One of the most important factors to consider in any social research is the positionality of the researcher—the background and position of the researcher in relation to the survey participants and the research setting. The need to consider the researcher's positionality and the extent of influence it could have on the research was important for two main reasons. Firstly, being a doctoral researcher and coming from a university abroad put me in the outsider position and could potentially influence the respondents' perception of me and the responses they provide. Secondly, having previously lived in the case study area for nearly two decades put me in the insider position where my knowledge of the local context could also be a source of bias especially in looking for predetermined responses to the survey questions. In order to account for the potential effect of my positionality on the reliability of the survey responses, a reasonable amount of time was spent at the beginning of the interviews to explain to the participants the purpose of the research, who is conducting the survey and how the data obtained will be used. Also, the interviewees were encouraged not to understate or overstate their responses to the survey questions since their responses would be anonymized but instead, to provide answers that reflect their circumstance as close as possible with respect to the content of the questionnaire.

Moreover, triangulation methods were employed to check the reliability of the responses to the questionnaire administered by the field assistants. The first step to ensuring reliability was achieved through the design of the questionnaire itself. Inbuilt triangulation included conditioning responses to some of the questions based on previous responses supplied by the interviewee. In addition to facilitating easy and quick administration of the questionnaire by removing non-applicable questions and answer options, designing the questionnaire this way allowed the researcher to determine if the right questions have been administered by the research assistants. Besides the design of the questionnaire, actual fieldworks were scheduled so that each selected proxy community was visited the same day by the research with a group of field assistants. At the end of each interview by a field assistant, a debriefing meeting was held to audit the completed questionnaire. Answers to the survey questions were then audited and interrogated. For example, given the employment type and educational background of a respondent, the reliability his or her self-reported income could be ascertained especially in cases of overestimation. This also ensured that research assistants could not make-up some of the answers along the way. Adopting this auditing approach also ensured that the research assistant could go back to fill in gaps in responses and ask for clarification where necessarily.

3.8 Data processing, analysis themes and statistical methods

Primary data obtained from households were edited after each day's interviews. The data were then collated, coded and entered in SPSS for further editing and processing for analysis. A plan to structure and report the findings from the survey was then prepared. This involved identifying appropriate statistical analysis methods for each of the analysis themes identified earlier in section 3.6.3. Table 3.7 shows the main research questions, the corresponding analytical themes and statistical analysis methods.

As indicated in table 3.7, basic descriptive statistics including measures of central tendency as well as inferential statistics including Principal Component Analysis, Correlations, standard Linear Regression and Logistic Regression models would be adopted to explore, understand and describe the data. An overview of the Principal Component Analysis, Linear Regression and Logistic Regression methods, the advanced statistical analysis methods to be adopted for the data analysis is presented in the sections that follow.

Table: 3.7: Household survey data analysis themes and statistical analysis methods

Research questions	Data analysis themes	Statistical analysis methods
1. What are the factors and underlying processes of residential location choice behaviour of heterogeneous households in urban areas?	▪ Analysis of background socio-demographic attributes of households	▪ Basic descriptive statistics (percentages, measures of central tendency) and parametric and non-parametric statistics
	▪ Examination of the housing market characteristics	▪ Linear regression, parametric and non-parametric statistics
	▪ Identification of the macro and meso-scale residential factor preferences of households.	▪ Principal Component Analysis, ▪ Kruskal-Wallis H Test of association and basic descriptive statistics
	▪ Examination of preferences for housing types.	▪ Multinomial Logistics Regression
	▪ Examination of preferences for housing tenancy types	▪ Multinomial Logistics Regression
	▪ Analysis of housing occupancy characteristics	▪ Basic descriptive statistics linear regression
	▪ Analysis of housing costs and affordability perceptions in the rental sector	▪ Basic descriptive statistics and non-parametric tests of association
2. What are the factors and underlying processes of job location choice behaviour of individual working members of the heterogeneous urban households?	▪ Examination of job market characteristics	▪ Basic descriptive statistics, parametric and non-parametric statistics
	▪ Analysis of the spatial distribution of employment	▪ GIS and descriptive statistics
	▪ Examination of the macro and meso scale job location factor preferences of individuals	▪ Principal Component analysis and basic descriptive statistics
	▪ Analysis of the determinants of Job location	▪ Binary Logistics Regression

Table: 3.7 continued: Household survey data analysis themes and statistical analysis methods

Research questions	Data analysis themes	Statistical analysis methods
3. What are the interdependence between the residential location choice and job location choice of the households	▪ Analysis of households' residential location change/mobility and reasons for change of residence, individuals' job location change, and reasons for job location change	▪ Basic descriptive statistics and cross tabulation analysis
4. What are the mobility patterns associated with the emergent residential-job location combinations?	▪ Analysis of home-work trip production and attraction among Traffic Analysis Zones (TAZ)	▪ GIS, descriptive statistics and cross tabulation analysis
	▪ Examination of the determinants of private car ownership	▪ Binary Logistic Regression
	▪ Examination of the determinants of active and motorised transport use	▪ Binary Logistic Regression
	▪ Examination of the determinants of choice between public transport modes	▪ Logistic Regression Models
	▪ Analysis of travel times and costs associated with home-work commuting and their determinants	▪ Descriptive analysis, correlation analysis and linear regression

3.8.1 Principal component analysis (PCA)

One of the major inferential statistical analysis method considered suitable for the analysis of the survey data is Principal Component Analysis (PCA). PCA is a multivariate statistical technique used to compute new variables from original variables by deriving linear correlations from combinations of the original variables: the new variables are called principal components (Abdi and Williams, 2010; Jollif and Cadinna, 2016).

Urban location choice and preferences, the phenomenon under investigation in this research is multi-faceted and cannot be measured directly using a single variable. Instead, several indicator variables were used to capture different facets of the phenomenon. In this research for example, 16 factors were used as indicators to capture the various dimensions of residential location preferences at the macro and meso scales while job location preferences were measures using 8 indicative factors. PCA becomes useful in the analysis of this data by extracting important information from the many factors and grouping them into a few constructs or components based on relationships between them. This helps to determine the underlying structure in the data while simplifying the description of the data by reducing the large number of factors into relatively small number. The PCA method is formalized in equation 3.1 as follows:

$$\boldsymbol{Y_i = b_i X_{1i} + b_2 X_{2i} + \cdots + b_n X_{ni}} \qquad \mathbf{3.2}$$

Where:

Y_i denotes component extracted from the data;

b denotes component loadings; and

X denotes the variables that cluster around the component

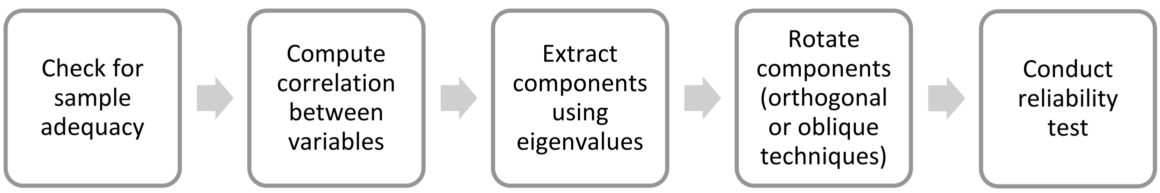


Figure: 3.9: Procedure for Principal Component Analysis. Source: adapted from Field (2013), page 684

PCA follows a set statistical analysis procedures involving checking for sample adequacy, computing the correlations between variables, extracting components based on eigenvalues, rotating the components and conducting tests of reliability on the results (see figure 3.9). Each of these steps are described below.

Generally, PCA requires large samples to make the results meaningful and reliable. For this reason, the adequacy of sample size for the analysis must be verified. Kasier-meyer-Olkin Measure of Sampling Adequacy (KMO) (Kaiser, 1970), is a technique used to check for sample adequacy for conducting PCA. KMO computes the ratio of the squared correlation between variables to the squared partial correlation between them. The resulting statistic varies between 0 and 1. As explained in Field (2013), a value of 0 indicates that the sum of partial correlations is large relative to the sum of correlations, indicating diffusion in the pattern of correlation. This implies that PCA is inappropriate. A value closer to 1 indicates that patterns of correlations are relatively compact and that PCA is appropriate. As a rule of thumb, KMO values greater than 0.5 indicate sampling adequacy for PCA.

Once sample adequacy is checked, the next step involves computing the Pearson correlations between each pair of variables in the data to obtain R-matrix estimates which transforms the original variables into linear components factors. Each component/factor has an associated factor loading expressed as a matrix (component matrix) in which columns represent each component and the rows represent loadings of each variable on each component.

The third stage of the PCA is component/factor extraction—deciding which components to retain and which ones to remove. Components are extracted using their associated eigenvalues[4]. Two complementary approaches are adopted: using scree plot which graphs the eigenvalues to visually depict their relative importance and using Kaiser's criterion (Kaiser, 1960). Kaiser's criterion is based on the idea that the eigenvalues represent the amount of variation explained by a component. Using this criterion, components/factors with eigenvalues greater than 1 are retained. Jollife (2002), however demonstrate that Kaiser's criterion is too restrictive and that components with eigenvalues greater than 0.7 could be retained.

[4] Eigenvalues are statistics generated as part of PCA which indicate the substantive importance of each of the components or factors; consequently, it is logical to retain components with higher eigenvalues.

Generally, most variables have high loadings on the most important component and small loadings on all other components, making interpretation difficult. For this reason, a technique called rotation is used to discriminate between components and to determine the degree to which variables load onto each component. Field (2013) identifies two main rotation techniques namely; Orthogonal rotation where components are rotated while keeping them independent or unrelated and oblique rotation where the components can correlate. The reliability of the extracted and rotated components is then tested.

3.8.2 Logistics regression model

Logistic regression analysis (LRA) is the second type of inferential statistical method adopted to analyse the survey data. LRA is the appropriate methods to measure the relationship between a discrete outcome variable that takes two or more possible values and one or more explanatory variables by estimating probabilities using a logistic function (Hosmer and Lemeshow, 2000; Faraway, 2016).

Categorical outcome[5] variables violate the assumption of linear relationship between variables held in simple regression analysis: LRA overcomes this violation by expression the simple linear regression equation in logarithmic terms, called the logit (Field, 2013). In its simplest form where the outcome variable has two categories, a binary logistic regression is use. The logistic equation takes the form:

$$\boldsymbol{P(Y) = \frac{1}{1 + e^{-(b_o + b_i X_{1i})}}} \qquad \textbf{3.3}$$

Where:

$P(Y)$ denotes the probability of an event Y occuring;

e denotes the base of the natural logarithms

b_0 denotes the constant

X_i denotes the predictor variable

b_1 denotes the co-efficient or weight attached to the predictor

[5] A Categorical variable could be binary or dichotomous meaning that it has two categories or polychotomous meaning that it has more than two categories.

Where the outcome variable has more than two categories, the logistic equation could be extended into a multinomial or polychotomous model to include the additional predictors X_2 and X_{ni} and their associated coefficients b_2 and b_n as depicted in equation 3.3

$$\boldsymbol{P(Y) = \frac{1}{1 + e^{-(b_o + b_i X_{1i} + b_2 X_{2i} + \cdots + b_n X_{ni})}}} \qquad \textbf{3.4}$$

Thus, LRA estimates the probability of an event occurring given known values of some explanatory factors. The resulting value from the equation ranges between 0 and 1, with a value close to 0 meaning that Y is very unlikely to have occurred, and value close to 1 meaning Y is very likely to have occurred. The associated parameters of the predictor variables in the equation are estimated from the sample data using maximum-likelihood estimates. This method selects coefficients that make the observed values most likely to have occurred (Kleinbaum and Klein, 2010).

LRA provide odds ratio estimates which is used to explain the relationship between the outcome and explanatory variables. The odds of an event occurring are defined as the probability of that event occurring divided by the probability of that event not occurring. If the value of the odds ratio greater than 1, then it indicates that as the explanatory variable increases, the odds of the outcome occurring also increases. Conversely, a value less than 1 indicates that as the predictor increases, the odds of the outcome occurring decreases (Field, 2013).

As outlined in table 3.7, the analysis of households' preferences for housing types (i.e. detached, semi-detached, compound and flat) and tenancy types (owner, renting and rent-free), as well as the determinants of individuals' job location (i.e. whether home-based or non-home) and travel mode choice involve either binary or polychromous outcome variables. LRA is considered the most suitable method of analysis for these cases of discrete outcome variables.

3.8.3 Multiple (Linear) regression model

Multiple regression analysis (MRA) is an inferential statistical analysis method used to explain and quantify the relationship between a scalar outcome variable and two or more independent/ explanatory variables using the method of ordinary least squares (Field, 2013). The MRA involves fitting a linear model to the data to produce values of the outcome variable from a set of explanatory variables that are selected based on sound theoretical rationale or previously proven causal relationships (Field, 2013). The linear model is specified in equation 3.4:

$$\boldsymbol{Y_i = (b_o + b_1X_{1i} + b_2X_{2i} + \cdots + b_nX_{ni}) + e_i} \qquad \textbf{3.5}$$

Where;

Y_i denotes the outcome or dependent variable;

b_0 denotes the intercept/constant, a parameter representing the value of the outcome variable when a predictor is zero;

$b_1 b_2 \ldots b_n$ the coefficient estimates of the predictor variables in the model;

$X_{1i} X_{2i} \ldots X_{ni}$ denotes the explanatory/predictor variables in the model; and

e_i denotes an unobserved error term

To assess how well a regression model specified fits the observed data, goodness of fit measures is used. R^2 , which is a fraction value between 0.0 and 1.0 provides a useful measure of how well a model fits and indicates the amount of explained variance in the model. This is obtained by dividing the sum of squares[6] for the model by the total sum of squares. The validity of any MRA depends on satisfying several assumptions. MRA assumes a linear relationship between outcome and explanatory variables, independence of residual terms, random, normal distribution of residuals, and equal variance of residuals (see Montgomery et al., 2015; Field, 2013).

As indicated in table 3.7, the analysis of the determinants of attributes of households such income, characteristics of housing market such as rent levels and room occupancy rates as well as the determinants of work travel times and transport costs all involve scalar outcome variables. The use of MRA was therefore considered suitable statistical analysis methods.

3.9 Reporting results of survey data

The analysis of the data is presented in the next two chapters—chapters four and five. Results of the data analysis addressing household's urban location choice (i.e. residential and job location choice) and the choice interdependence is presented first in chapter four while that of

[6] The sum of squares is derived by calculating the average of scores in the data and finding the difference between each score and the average, squaring each of these numbers and adding the squared numbers.

mobility characteristics associated with the residential-job locations of individual adult working members within household is presented in chapter five.

3.10 Chapter Summary

This chapter has set out the overall approach and methodology to empirically examine the residential location choice behaviour of households, the job location choice behaviour of adult individuals in the household, the interdependence between residential and job location choice, and the mobility characteristics associated with individuals' residential and job location combinations.

The case study design, which allows for detailed investigation of a phenomenon within a specific context was adopted as the most appropriate approach to examine the urban location choice and mobility patterns nexus. Following from this and applying a set of case study selection criteria, the Kumasi metropolis in Ghana, West Africa was presented as the case study area for this research.

A comprehensive program to obtain the relevant data from the case study area was put forward. This involved identifying the specific study variables from the empirical research questions. The study variables were then translated into a structured research questionnaire that was used to elicit data from households and individuals regarding their location choice decision-making processes and daily travel characteristics. The analytical themes from the survey data and the appropriate statistical methods that would be adopted for the analysis were then outlined and discussed.

The next two chapters immediately following this chapter (i.e. chapters four and five), will present the analysis of the survey data. Analysis of households' residential and job location choice will be presented first in chapter four followed by the analysis of mobility characteristics associated with the home-work location pairs of individuals in the household in chapter five.

CHAPTER FOUR: UNDERSTANDING RESIDENTIAL AND JOB LOCATION PREFERENCES AND CHOICE INTERDEPENDENCE—AN EMPIRICAL ANALYSIS

4.1 Introduction

Residential and job locations constitute the long-term choice of households and individuals in urban areas. The observed distributions of home locations and employment locations are the outcome of the interplay of several factors which interact at different scales. At the level of households and individuals, factors including life-cycle stage, household structure and composition, family relations and social networks, and income, shape heterogeneous preferences for the place of residence. Location-defining attributes including opportunities to realize reasonable home-work distance separation, availability and access to essential amenities and character of neighbourhoods, interact with socio-demographic factors to shape emergent residential locations at the meso-scale. At a more micro-level of location choice, characteristics of dwellings including dwelling type, size, tenancy and price also interact with socio-demographic attributes to determine residential location outcome. The job locations of individuals within the household are also influenced by interaction between personal-level characteristics such as levels of educational attainment and skill levels of job seekers, wider circumstances of the job seeker's household, job market factors including job availability, opportunity for higher wages, and urban structural factors such as the spatial distribution of job centres relative to other activity locations in the urban area.

The place of residence and work-place are connected, suggesting at least conceptually, that both locations are responsive to each other over different time horizons. For analytical purposes, the interdependence between these two choice sets has been modelled in some cases assuming a conditional choice process in which the place of residence is assumed to come first followed by the place of work or vice versa. Other analytical approaches have also adopted a conjoint approach in which choice for both the place of residence and work-place are assumed to occur at the same time. The validity of any of these assumptions are often context specific. Moreover, even within a given context, different households and individuals may decide their locations implementing any of the above assumptions. The problem becomes even more complex when residential and job relocations are introduced into the mix. Thus, empirically examining the interdependence between these two long-term choice sets is critical to a better understanding and representation of the choice process within any given context.

The focus of this chapter is to address the location choice aspect of the empirical research objective of this thesis. As articulated in Chapter three, the first objective of this thesis is to understand empirically, the residential and job location choice of households and the associated mobility patterns in the Kumasi Metropolis, the case study area. This chapter, being the first of two empirical studies conducted to address the above objective, presents analyses of the urban location choice process of households and individuals, focusing on residential location choice and job location choice and the interdependence between the two choice sets in the case study metropolis. To this end, the analysis addresses the following specific questions:

i. What are the factors and underlying processes of residential location choice of heterogeneous households in the Kumasi metropolis?
ii. What are the factors and underlying processes of job location choice of individual working members of the households in the metropolis? and
iii. What are the interdependence between the residential location choice and job location choice process?

4.2 Overview of the data and statistical analysis methods

In Chapter three, a detailed discussion of the overall methodology covering the empirical analysis of urban location choice was presented. Details of the sampling approach, the study variables, the survey questions, data analysis themes and statistical methods to be adopted were also presented in Chapter three. In view of this, only a summary of the data and statistical methods used for the analysis in this chapter is presented in this section.

The empirical analysis of residential and job location choice is based on primary data obtained from households and individuals in the Kumasi metropolis through cross-sectional survey. A total of 665 randomly sampled households from 32 locations across the metropolis were interviewed to elicit retrospectively, the factors they considered in deciding their places of residence. From this sample of households, a total of 1,158 individual adults working members were also interviewed to obtain data on their job location choice. The fundamental underlying assumption of this study is that residential location decisions are taken at the level of the household while job location decisions are taken by adult working individuals considering circumstances of the wider household to which they belong.

Relevant statistical analysis methods ranging from simple descriptive statistics to more advance multivariate methods such as Linear Regression (LR), Binomial Logistic Regression (BLR), Multinomial Logistic Regression (MLR) and Principal Component Analysis (PCA) are employed in the data analysis. Detailed discussion of the LR, MLR and PCA methods including their mathematical representations and assumptions can be found in Chapter three.

The determinants of dwelling type and tenancy choice as micro-scale residential choice alternatives are examined using the MLR method. At the meso-scale, basic descriptive statistics and a logistic regression model is used to examine the determinants residential location among the three-broad urban-zones in the metropolis namely; historical-core, inner-suburb and outer suburb. Descriptive statistics are also applied to examine housing occupancy characteristics as well as the rent-to-income affordability thresholds of households in the rental market.

As described under the research instrument design and data collection sections of the research methodology, respondents were given a set of factors covering their residential and job locations two which they indicated the importance they attach to each of them on a five-point Likert scale. In total, households evaluated 16 items in relation to their residential location choice while individual workers evaluated 8 factors in relation to their job location choice. Results of this data, which constitute the meso-scale choice considerations of households and individuals with respect to the two choice processes of interest are presented using the PCA method. As a multivariate statistical technique, the PCA method distils the various factors into a few constructs or components using eigenvalues determined by the statistical relationships between the individual factors; for each extracted component, the percentage of variance explained are also indicated (Abdi and Williams, 2010; Field, 2013). Moreover, the determinants of choice between home-based and non-home-based employment locations of individual workers are examined by fitting a BLR model to the data. Finally, descriptive statistics of mainly percentages and measures of central tendencies are used to present results of the data elicited on the choice interdependence between the place of work and place of residence.

4.3 Chapter organization

Results of the data analysis follows in seven interrelated sections. In the first section, the background characteristics differentiating the households and individuals whose location choice preferences are being analysed is presented. This is followed with a discussion of the housing market characteristics summarizing the characteristics of dwellings as the households' choice alternatives. In the third section, the determinants of residential location choice at the macro and meso scales of location are extracted. Preferences for dwelling types and tenancy that determines choice at the micro-scale as well as occupancy characteristics and housing affordability thresholds are examined in the fourth section. Next, characteristics of the metropolitan job market and the determinants of employment locations are discussed. The penultimate section examines the choice interdependence between individuals' place of residence and place of work followed by a summary of the results and key findings in the concluding section.

4.4 Background socio-demographic characteristics of households

Socio-demographic characteristics reflect diversity in the population, which in turn, influence preferences for different combinations of urban locational attributes. In view of this, the household survey elicited data on the coupling rate, family size, age distribution and income distribution among households in the Kumasi metropolis.

Table 4.1 provides a summary of attributes defining households in the metropolis. Marital status as dimension of life-cycle stage was 65% among households while single and co-habiting households constituted 16% and 15% respectively. The average family size in the sample was 4 persons (SD = 1.788), compared to the metropolitan average family size of 5 persons reported by the 2010 population and housing census. The average age of heads of households interviewed was 43 years (SD = 12. 622) while that of children/dependents was 14 years (SD = 14.00). In terms of levels of educational attainments of household heads, the analysis found that about 41% of household heads and 56% of co-head, in the case of couples had only basic level of education. About 29 % and 6% of household-heads and co-heads were tertiary-educated. Using qualifications as a proxy of skill levels, it can be concluded that most of household heads and their co-heads had low to intermediate levels of skill.

Table 4.1: Summary of socio-demographic characteristics of households

Category		Attribute	Frequency	Percentage
a.	Marital status	Single (never married)	109	16
		Couple (Married)	432	65
		Couple (Co-habitation)	27	4
		Single (Divorced & Widowed)	97	15
		Total	**665**	100
b.	Age of household heads	18-24	17	3
		25-34	160	24
		35-44	204	30
		45-54	153	23
		55-64	87	13
		65+	44	7
		Total	**665**	**100**
c.	Age of Children/Dependents	>5	207	14
		5-9	269	19
		10-14	310	22
		15-19	276	19
		20-24	212	15
		25-29	115	8
		30-34	36	3
		35-39	14	1
		Total	**1440**	**100**
d.	Educational attainment of household-heads	Basic School	273	41
		Secondary & Post-Secondary	199	30
		Tertiary	193	29
		Total	**665**	**100**
e.	Educational attainment of partners of household-heads	Basic School	257	41
		Secondary & Post-Secondary	174	30
		Tertiary	28	29
		Total	**459**	**100**

Source: Based on Field Survey, February 2015

The households interviewed were also differentiated by income levels. Out of the 665 households, income data was obtained from 643 of them, representing 97% of all households surveyed. On the average these households had 2 adult-working members (SD = 0.737). The distribution of household income was asymmetrical (Skewness = 3.02; Kurtosis = 16.86.; SD = 1327.4799), implying that the skewness was substantial and that the distribution was far from symmetrical.

Total family income was taken as the sum of reported monthly earnings of all working members within the household aged 15 years and above. An alternative method would have been to use income per capita by dividing total household earnings by family size. Background analysis of the data, however, showed that occupation type, educational attainments and number of working members all correlated highly with income levels. In addition, most households (i.e. 70%) who also happen to be couples had two working members whose educational levels and occupation type, as determinants of income, were found to be similar.

It was therefore concluded on the basis of this information that within the context of the study area, total household income as summation of individual working members' earnings reflect closely the different levels of income, ranging from low-income to high-income. Indeed, as the analysis will later show, the income categories derived from the data corresponds to various indicators of socio-economic status including dwelling types choice (i.e. compound housing vs. detached, semi-detached and flats in multi-storey buildings), housing tenancy choice (i.e. owner-occupier vs. renting and rent-free), employment location (i.e. home-based vs. non-home-based) and transport mode choice (i.e. car-ownership vs. public transport and walking).

Households were categorized into different groups based on their earnings. Pre-existing data on income categories to which the survey data could be conformed does not exist in Ghana. Instead, a national absolute poverty line of GH¢1,314.00 per annum or GH¢109 per month, has been established (Ghana Living Standards Survey, Round 6, 2014). For this reason, the income data obtained through the survey was first divided into percentiles. The average monthly income of households was GH¢1,200.00 (US$312), with lowest (5th percentile) and the highest (100th percentile) incomes of GH¢300.00 (US$ 78) and GH¢ 14,000 (US$ 3,645) respectively. Households were classified into six income groups based on the percentile distributions, using the basic monthly income of GH¢109 (US$28) as the baseline as shown table 4.2.

Table 4.2: Six classes of households based on total monthly earnings[7]

	Monthly Income Range		Descriptive Statistics	
Income group	**GH¢**	**US$**	**Percentage**	**Percentiles**
Urban poor	below 109.5	28.5	1	1st
Low	150 -700	39 - 182	25	2nd - 25th
Lower-middle	750 - 1200	195 -312	27	26th - 50th
Upper-middle	1250 - 2000	325 - 520	23	53rd -75th
High	2050 - 4000	533 - 1040	20	76th - 95th
Rich	4100 - 14000	1066 - 3640	4	96th - 100th

Source: Based on Field Survey, February 2015

Moreover, the relationships between the income-groupings and attributes of the households including the number of working persons, educational attainment of workers, coupling rate and family size are examined. A summary of the results is presented in figure 4.1. The analysis show an association between total household income and the number of working members in the household ($r = 0.55$ $p < 0.001$).

[7] Dollar equivalent of incomes computed using the prevailing rate of US$1 to GH¢ 3.84 as of January 2016

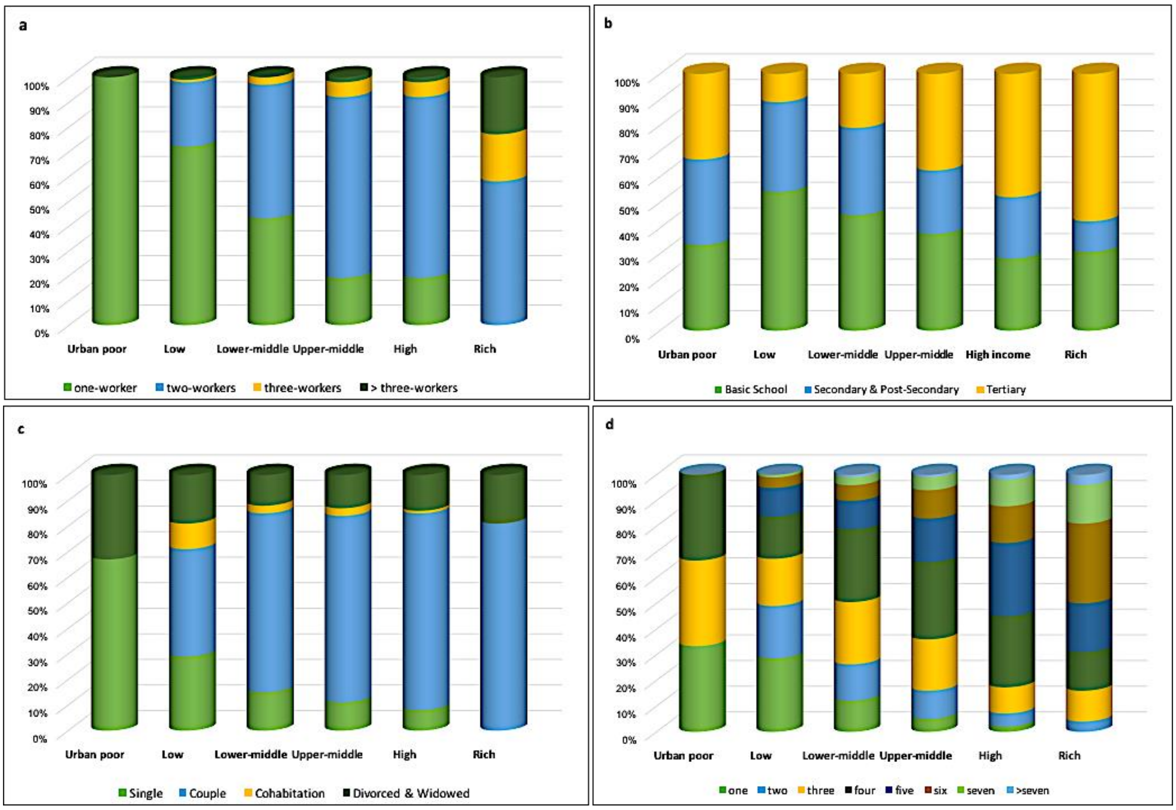


Figure 4.1: Relationships between income-groupings of households and (a) number of adult-working members (b) levels of educational attainment of household-heads (c) coupling status and (d) family size.

Fitting a linear regression model in which levels of educational attainment of household heads and their co-heads, as well as the number of working members were set as explanatory variables of total household income show that as educational attainment of household heads and co-heads increased, total household income increases by GH¢ 313.822 ($p < 0.001$) and by GH¢ 99.290 ($p < 0.001$) respectively, controlling for the number of working members. Also, any additional working person in the household increases total household income by GH¢ 713.362 holding educational attainment of household heads and co-heads constant ($p < 0.001$). Together, these factors accounted for 26% of the variance in household incomes. Moreover, across all income groups, coupling rate was 70% or higher except for low-income earners among whom 42% were married (see figure 4.1c.) Also, family size appears to be higher (five person) among households classified as high income and rich and lower (three persons on the average) among urban poor, low and lower income households compared to the average family size of 4 among all households.

4.5 Housing characteristics

Analysis of the housing market characteristics provides an important background information about the context within which households realize their residential location and housing preferences. In view of this, the survey elicited data on the types and size of dwellings as well as tenancy types in the metropolis.

Four unique types of dwellings were identified in the Kumasi metropolis from the survey data. As shown in figure 4.2a, nearly half (49%) of all dwellings were of the traditional compound housing[8] type. This housing type remains the dominant dwelling occupied by households in the in the historical-core, inner-suburban and outer-suburban zones of the metropolis. Notwithstanding, as depicted in figure 4.2b, the proportion of compound dwellings decrease as distance increases farther away for the historical-core. Generally, as one moved away from the inner-core locations of the metropolis, the proportion of single family housing units (i.e. detached and semi-detached housing) tends to increase.

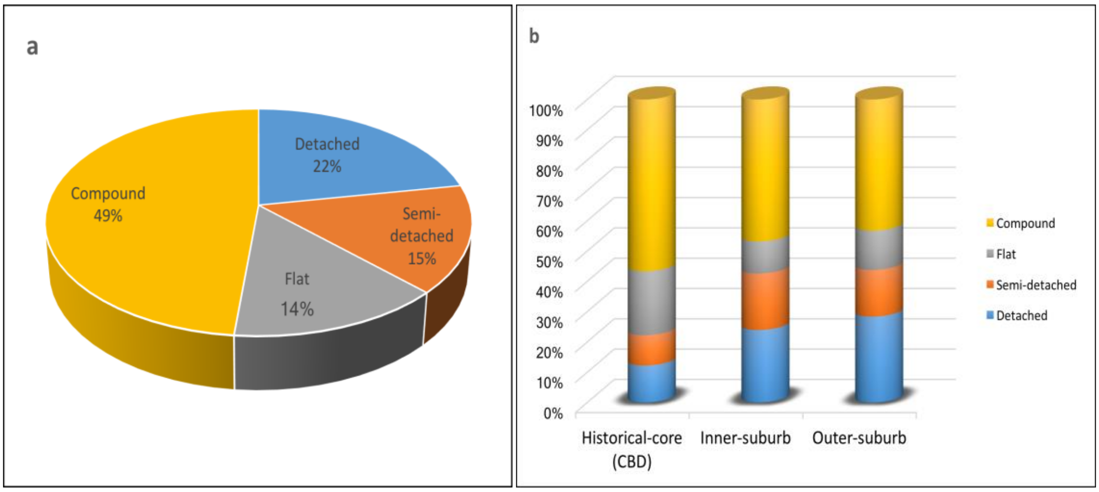


Figure 4.2: (a) Dwelling types of households (b) distribution of households' dwellings among the three urban zones. Source: Based on Field Survey, February 2015

Dwelling size is another dwelling-defining attribute. On the average, detached housing had 4 bedrooms; semi-detached and flats and multi-storey buildings also had 2-bedrooms on the average. Compound housing, built for multiple occupation, had on the average, 11 rooms per house (SD = 6.14).

[8] The compound house is one-storey structure with a square or rectangular open courtyard surrounded by a series of single room units often for multiple-habitation (Sinai, 2001).

Besides the type and size of dwellings, the survey also asked households to indicate their tenure arrangements for their current dwellings. Whereas 16% of tenancies were owner-occupied, about 34% of households lived rent-free.

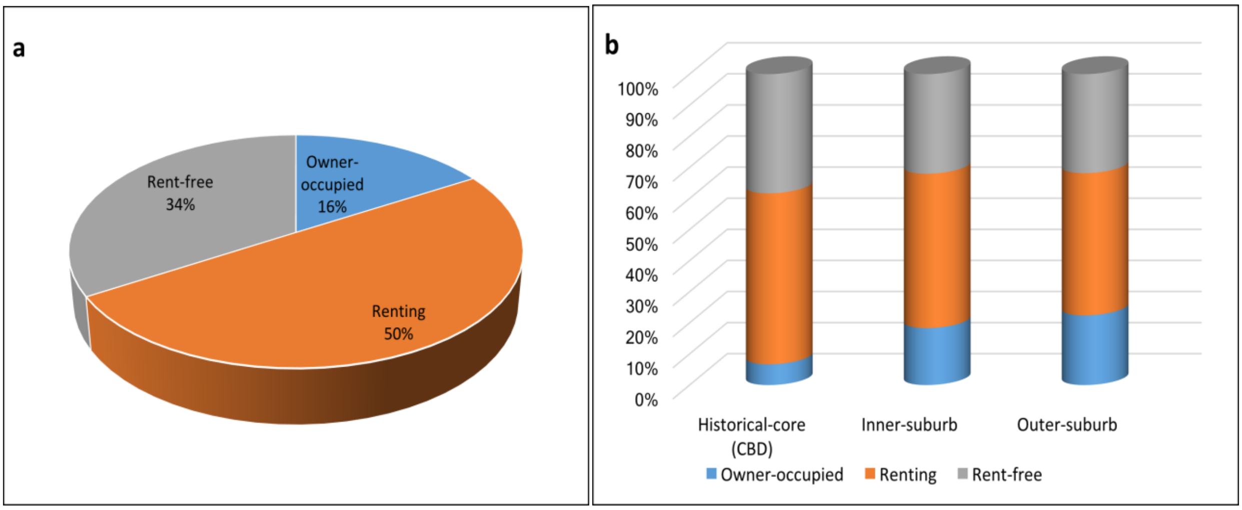


Figure 4.3: (a) Tenancy types of households (b) distribution of housing tenancies among the three urban zones.
Source: Based on Field Survey, February 2015

Using Chi-square statistics, the analysis found positive associations between tenancy types and the three urban zones of the metropolis (see table 4.3).

Table 4.3: Relationship between dwelling types and tenancy arrangements in the urban-zones

Urban zones	**Dwelling Types**	**Tenancy Arrangements (Percentages)**			
		Owner-occupied	Renting	Rent-free	Total
Historical-core	Detached	27	45	27	100
*(Fisher's exact test = 60.650**	Semi-detached	28	72	0	100
Cramer's V = 0.385)*	Flat	3	58	39	100
	Compound	0	53	47	100
Inner-suburb	Detached	49	21	31	100
(*Fisher's exact test = 105. 879**	Semi-detached	19	51	30	100
Cramer's V = 0.385)*	Flat	20	67	13	100
	Compound	2	60	38	100
Outer-suburb	Detached	54	25	21	100
*(Fisher's exact test = 58. 401**	Semi-detached	13	58	29	100
Cramer's V = 0.318)*	Flat	12	65	23	100
	Compound	8	49	43	100

Note: *$p < 0.001$

Among all the zones, rent-free housing was dominant within compound housing and flats in multi-storey buildings. Rent-free tenancy arrangement is characterized by independent households, belonging to an extended family, living together in a compound house owned by members of the kinsfolk without paying any rent (see e.g. Tipple et al., 1997; Acheampong, 2016). In the suburban and peri-urban areas, rent-free housing is also provided under what is known as *'caretaker'* arrangement where owners of uncompleted houses engage a caretaker to look after the house and building materials and keep the site tidy; in return, the caretaker does not pay rent (see e.g. Gough and Yankson, 2011). Furthermore, renting was concentrated in compound houses and flats whilst house ownership was common among households who lived in detached and semi-detached housing within the inner and outer sub-urban zones.

4.6 Analysis of residential location choice: An examination of macro and meso level choice factors

The foregoing sections have presented the relevant background socio-demographic information based on which different households and their characteristics have been identified. This section begins the analysis of residential location preferences of households by examining the interaction between the households' characteristics and preference for various macro-scale and meso-scale attributes that informed the choice of their current residential locations.

4.6.1 Residential location distributions in the urban zones

Out of the 665 households interviewed, 180 (27%) had their homes located in the historical-core neighbourhoods of the metropolis. The remaining 284 (43%) and 201 (30%) lived in neighbourhoods located within the inner-suburban and outer-suburban zones of the metropolis respectively. Figure 4.4 shows the distribution of households of different income groups among the three urban-zones. Among households living in the historical core of the metropolis, the majority (35%) and (29%) fall into the low-income and lower-middle-income categories respectively. The proportion of low-income households decreases to 24% and 18% within the inner-suburb and outer-suburbs respectively. Moreover, the proportion of lower-middle income households decreases to 25% within the inner-suburb but increases to 29% at the outer-suburb. Among, upper-middle income and high-income households, their percentage distribution increases progressively from the historical core to the outer-suburb. The proportion of rich households remains constant at 4% within the historical core, inner-suburb and outer-suburban areas.

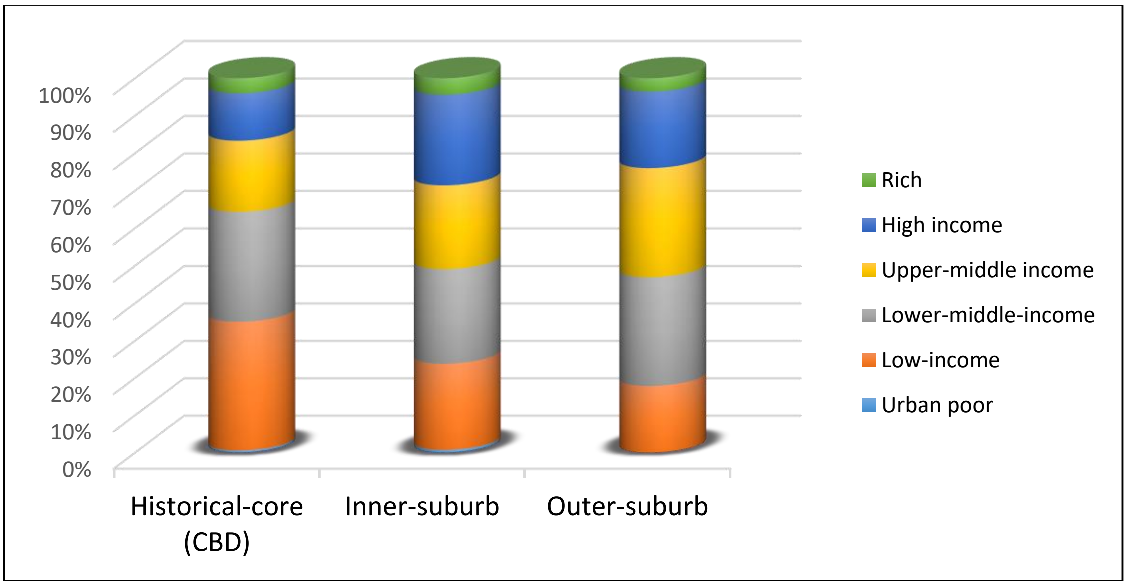


Figure 4.4: Distribution of households of different income groups among the three urban zones.
Source: Based on Field Survey, February 2015

Chi-square analysis revealed that the income group of households has a significant effect on the zone of residence[9] (Chi-square = 34.472, df = 4, $p < 0.001$), although the effect size was small (Cramer's V statistic = 0.12 $p < 0.001$). Odds ratio estimates also shown that the likelihood of low-income households taking residence in the historical-core was 2.5 times higher than households of relatively higher incomes (Odds ratio = 2.429 Wald = 21.082, df = 1, $p < 0.001$).

4.6.2 Extracting residential location choice factors: A principal axis factoring method

Households gave a retrospective account of the factors that they considered in choosing their current residential locations. They evaluated on a 5-point Likert scale from 1- "*very important*" to 5- "*not-important-at-all*", 16 location choice factors that were presented to them. An outline and descriptive statistics of the factors evaluated by the households are presented in table 4.4

[9] To meet the assumption that zero cells (0%) have expected counts less than 5 in the contingency table associated with the chi-square statistics (Field, 2013), the income group data was recoded such that households classified as urban poor and low income were collapse into one category.

Table 4.4: Descriptive analysis of 16 evaluation items presented to households

Preference Items	**Mean**	**Std. Deviation**	**N**
Proximity to my workplace	2.24	1.175	665
Proximity to workplace of my partner/spouse	2.92	1.225	665
Prestige of the neighbourhood	2.81	1.264	665
Proximity to major road	2.04	1.034	665
Proximity to transport station/terminal	2.13	1.083	665
Proximity to essential services (e.g. banking, postal, health)	2.25	1.087	665
Proximity to markets/shopping centres	2.19	1.060	665
Proximity to school(s) for my children	2.65	1.173	665
Availability of opportunities for better paid jobs and businesses	2.49	1.200	665
Quiet and serene neighbourhood	2.47	1.248	665
Less traffic and noise pollution	2.44	1.211	665
Safety for my family	1.71	0.888	665
Preference to live closer to friends	3.39	1.177	665
Preference to live closer to family/relatives	3.30	1.275	665
Preference to live closer to people of same ethnicity	3.42	1.271	665
Preference to live among people I think are of the same socio-economic status as me	3.42	1.224	665

Source: Based on Field Survey, February 2015

A Kruskal-Wallis Test was conducted to ascertain the level of importance households within the six income groups identified attached to each of the residential location items. The test ranks, from highest to lowest, the mean score of each income group on each of the evaluation items. Thus, the group with the lowest mean rank is the group with the greatest number of lower scores in it while the group with the highest mean rank have the greater number of high scores within it (Field, 2013).

Out of the 16 items, eight yielded statistically significant results **(**see table 4.5). The eight items referred generally to amenity and essential services proximity, closeness to relatives and people of the same ethnicity, and locations closer to work place of adults and schools of children. Consistently across all 8 items evaluated by the households, households belonging to urban poor and low-income category scored the highest mean, indicating for example, that they attached higher importance to proximity to transport infrastructure (i.e. roads and terminals), essential services, markets and shopping centres, among others, than the other households.

Table 4.5: Kruskal-Wallis H Test of association between residential location factors and household incomes

Items	Kruskal-Wallis H Test Mean Ranks for Income Groups						Chi-Square
	Urban poor	Low	Lower-middle	Upper-middle	High	Rich	
Proximity to major roads	570.67	358.28	313.64	311.19	302.27	289.69	16.899*
Proximity to terminals	580.00	362.83	298.92	318.83	292.98	362.50	23.941**
Proximity essential services	462.67	359.08	304.10	321.26	296.89	330.23	14.121*
Proximity to markets/shopping centres	512.83	347.4	312.33	316.71	290.36	397.92	17.456*
Proximity to school(s) for my children	519.33	369.05	314.79	292.63	302.14	330.73	20.332**
Proximity to workplace of spouse	429.33	366.71	320.84	296.50	295.68	312.85	16.656*
Live closer to family/relatives	464.67	282.39	303.94	345.15	348.09	406.29	22.890**
Live closer to people of same ethnicity	359.00	307.86	298.57	329.22	342.59	419.87	14.078**

Note: *P <0.05, **P <0.001

To extract the most important residential location factors from the 16 items across all household, a principal axis factor analysis was conducted. Firstly, the Kaiser-Meyer-Olkin Measure (KMO) of the factor analysis verified the sampling adequacy for the analysis with KMO value of 0.825, which is well above the acceptable limit of 0.5. Secondly, Bartlett's test of sphericity was significant (approximate $X^2 = 3697.813$, df = 120, $p < 0.001$), and the KMO values for each of the 16 items produced by the anti-image correlation matrix were greater than 0.71, which were well above the acceptable limit of 0.5 (see Field, 2013). Finally, the determinant of the correlation matrix of 0.004, was bigger than the cut-off value of 0.00001 indicating that there was no problem of multi-collinearity among the variables.

Initial analysis was run to obtain eigenvalues for each of factor in the data. As shown in table 4.6, four factors obtained eigenvalues over Kaiser's criterion of 1, and together, explained 59.8% of the variance. Using the scree-plot and Kaiser's criterion for eigenvalues, as well as given the large sample size (i.e. N= 665), the four factors were retained.

Table 4.6: Initial Eigenvalues estimates of principal factors analysis on 16 residential location factors

Factor	Initial Eigenvalues Total	Initial Eigenvalues % of Variance	Initial Eigenvalues Cumulative %	Extraction Sums of Squared Loadings Total	Extraction Sums of Squared Loadings % of Variance	Extraction Sums of Squared Loadings Cumulative %	Rotation Sums of Squared Loadings Total
1	4.909	30.682	30.682	4.400	27.503	27.503	3.471
2	2.153	13.455	44.138	1.746	10.915	38.418	2.723
3	1.453	9.080	53.218	1.049	6.557	44.975	2.547
4	1.048	6.553	59.770	.517	3.228	48.203	3.077
5	.866	5.412	65.182				
6	.781	4.884	70.067				
7	.734	4.584	74.651				
8	.678	4.239	78.890				
9	.588	3.675	82.565				
10	.544	3.400	85.965				
11	.527	3.294	89.259				
12	.466	2.910	92.169				
13	.417	2.605	94.774				
14	.344	2.148	96.922				
15	.293	1.834	98.756				
16	.199	1.244	100.000				

Next, the oblique rotation method (direct oblimin) which assumes that the extracted factors are related rather than independent (Field, 2013), was used to extract factors. Tables 4.7 and 4.8 show the factors loadings and factor correlation matrix after rotation of the factors respectively.

Table 4.7: Factor loadings based on oblique rotation method

Preference Items	Factor 1	Factor 2	Factor 3	Factor 4
Proximity to transport station/terminal	**.833**			
Proximity to markets/shopping centres	**.734**			
Proximity to essential services (e.g. banking, postal, health)	**.726**			
Proximity to major road	**.640**			
Preference to live closer to people of same ethnicity		**.782**		
Preference to live among people I think are of the same socio-economic status as me		**.677**		
Preference to live closer to friends		**.662**		
Preference to live closer to family/relatives		**.639**		
Quiet and serene neighbourhood			**.914**	
Less traffic and noise pollution			**.907**	
Safety for my family			.296	.248
Proximity to workplace of my partner/spouse				**.679**
Proximity to school(s) for my children				**.582**
Proximity to my workplace	.118			**.519**
Availability of opportunities for better paid jobs and businesses	.198	.172		.252
Prestige of the neighbourhood		.210	.223	.230
Eigenvalues	4.91	2.15	1.453	1.048
% of Variance	30.7	13.51	9.10	6.60

Note: Absolute co-efficient values below 0.1 have been suppressed in the table.
Factor loadings over 0.40 appear in bold.

Table 4.8: Factor correlation matrix from oblique rotation method

Factor	1	2	3	4
1	1.000	0.356	0.380	0.570
2	0.356	1.000	0.076	0.422
3	0.380	0.076	1.000	0.450
4	0.570	0.422	0.450	1.000

As show in table 4.7, four items clustered around factor one, namely proximity to transport terminal, proximity to market and shopping, proximity to essential services and proximity to roads. Factor one therefore represents the construct '*proximity to major infrastructure and amenities'* as a residential location choice consideration**.** This factor accounted for 30.7% of the total variance in residential location explained by the factor analysis. The corresponding eigenvalue is 2.15.

The second factor, represents '*family ties and social networks'.* The presence of this factor implies that households placed premium on the need to live closer to family and friends. The fact that most households live in traditional compound houses together with relatives of the extended family in the Kumasi metropolis lends credence to this. In addition, factor two suggest evidence of the need by households to segregate along ethnic lines in their places of residence. Indeed, most the population in the metropolis (75%) are Akans, the dominant ethnic group in Ghana. Even so, there is a clear pattern of residential segregation along ethnic groups in the metropolis, particularly among ethnic minority groups in neighbourhoods such as Fante Newtown, Dakodwom, Anloga and Ayigya. '*Family ties and social networks'* as a construct, accounted for 13.5% of the variance in residential location choice among the households interviewed. The associated eigenvalue is 2.15.

The third factor extracted from the analysis reflect '*character of neighbourhood'*. On this factor, two aspects of locational attributes relating to traffic volumes and flow as well as the general quietness and serenity of neighbourhoods were considered important to the households. This factor explained 9.1% of the variance in residential location choice. The corresponding eigenvalue is 1.453.

Finally, three items evaluated by the household namely; proximity to workplace of working members and to the school(s) of children in the households clustered around the fourth factor. This last factor therefore encapsulates the '*core activities'* (i.e. work-place and school location) of the household members. It is not surprising that workplace and school proximity consideration would cluster as one factor because these are the two major activities in which adult members and children within the household spend most of their time daily. Proximity of the place of residence to these core activity locations accounted for 6.6% of the variance in residential location choice, with a corresponding eigenvalue of 1.048.

4.7 Analysis of housing choice: an emphasis on preferences for attributes of dwellings

Having established the macro and meso level locational factors considered important by households in the previous section, this section examines how preferences for dwelling types and tenancy vary among the different households in the case study metropolis.

4.7.1 Examining preferences for housing types: a multinomial logistics regression

A hierarchical multinomial logistic regression model was specified to ascertain the relationship between socio-demographic attributes of households and their current dwelling types. A summary of the variables used in the regression is presented in table 4.9.

Table 4.9: Variables in logistics regression analysis of dwelling types preferences

	Variable name	**Type and coding**
Dependent variable	Dwelling types	Categorical, coded: 1- detached; 2-semi-detached; 3-flat; and 4-compound
Independent variables	Income-group	Categorical coded: 1-low; 2-lower-middle; 3-upper-middle; 4-high; and 5-Rich
	Marital status	Categorical coded: 1-single; 2-married; 3-co-habiting; and 4-single (divorced or widowed)
	Family size	Scale variable
	Educational attainment of household heads	Categorical coded: 1- Basic School; 2- Secondary & Post-Secondary; and 3- Tertiary
	Residential zone of residence	Categorical coded: 1- Historical-Core; 2- Inner-suburb; and 3-Outer-suburb

The model fitting information indicated the final model was a better fit than the original model (X^2 = 163.233, df = 39, p < 0.001) while the Goodness-of-fit measures showed that the model is a good fit to the data (Pearson statistics: Chi-square = 1017.030, df = 963, p = 0.110; and Deviance statistics: X^2 = 879.079, df = 963, p = 0.975).

Given that compound housing represents nearly half (48%) of all dwelling units represented in the total sample, this was specified as the reference category for the dependent variable. Thus, the model is used to determine and predict the odds/likelihood of a household of some defining socio-demographic attributes living either in a detached, semi-detached and flat as compared to living in a compound house. Results of the MLR analysis is presented in table 4.10

Table 4.10: Logistics regression analysis of the likelihood of dwelling type choice by households

		95% Confidence Interval for Exp(B)		
Detached vs. Compound	*b(SE)*	Lower Bound	Odds Ratio	Upper Bound
Intercept	2.448 (0.829) *			
Low-income	-1.969(0.757) *	.032	0.140	.616
Lower-middle	-1.648(0.729) *	0.046	0.192	.803
Upper-middle	-1.800(0.727) *	0.040	0.165	.687
High	-1.384(0.731) *	0.060	0.251	1.049
Rich	0^b			
Single (never married)	-.825(0.429) *	0.189	0.438	1.016
Couple (married)	-.184(0.320)	0.444	0.832	1.557
Couple (co-habiting)	.401(0.580)	0.479	1.493	4.657
Single (divorced and widowed)	0^b			
Basic School	-1.805(0.293) ***	0.093	0.164	.292
Secondary & Post-Secondary	-.978(0.283) **	0.216	0.376	.655
Tertiary	0^b			
Family size	.011(0.077)	0.869	1.011	1.177
Historical-Core	-1.093(0.314) **	0.181	0.335	.621
Inner-suburb	-.308(0.246)	0.453	0.735	1.190
Outer-suburb	0^b			
Semi-detached vs. Compound				
Intercept	.526 (0.950)	0.023		
Low-income	-2.149(0.821) *	0.032	0.117	0.583
Lower-middle	-1.918(0.784) *	0.044	0.147	0.683
Upper-middle	-1.619 (0.772) *	0.097	0.198	0.899
High	-.828(0.767)		0.437	1.966
Rich	0^b	0.391		
Single (never married)	.232(0.598)	1.075	1.261	4.070
Couple (married)	.996(0.471) *	0.870	2.707	6.821
Couple (co-habiting)	1.358(0.764)		3.887	17.362
Single (divorced and widowed)	0^b	0.174		
Basic School	-1.136(0.311) ***	0.178	0.321	0.591
Secondary & Post-Secondary	-1.069(0.335) **		0.343	0.662
Tertiary	0^b	0.851		
Family size	.004(0.085)	0.306	1.004	1.186
Historical-Core	-.495(0.351)	0.599	0.609	1.212
Inner-suburb	.049(0.286)		1.050	1.840
Outer-suburb	0^b			

Note: R^2 = 0.224 (Cox and Shell), 0.245 (Nagelkerke), *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$,
[b] This parameter is set to zero because it is redundant

Table 4.10 continued: Logistics regression analysis of the likelihood of dwelling type choice by households

		95% Confidence Interval for Exp(B)		
Flat vs. Compound	*b(SE)*	Lower Bound	Odds Ratio	Upper Bound
Intercept	1.071 (0.878)			
Low-income	-2.043 (0.776) *	0.028	0.130	0.594
Lower-middle	-2.470(0.772) **	0.019	0.085	0.384
Upper-middle	-1.739(0.747) *	0.041	0.176	0.760
High	-1.301(0.752)	0.062	0.272	1.187
Rich	0^b			
Single (never married)	-.099(0.434)	0.387	0.906	2.121
Couple (married)	-.570(0.343)	0.289	0.566	1.107
Couple (co-habiting)	.070(0.734)	0.255	1.073	4.520
Single (divorced and widowed)	0^b			
Basic School	-.967(0.332) *	0.198	0.380	0.729
Secondary & Post-Secondary	-.671(0.346) *	0.260	0.511	1.007
Tertiary	0^b			
Family size	.146(0.080)	0.989	1.158	1.355
Historical-Core	.280(0.317)	0.710	1.324	2.466
Inner-suburb	-.418(0.325)	0.349	.659	1.244
Outer-suburb	0^b			

Note: R^2 = 0.224 (Cox and Shell), 0.245 (Nagelkerke), *p < 0.05, **p < 0.01, ***p < 0.001,
[b] This parameter is set to zero because it is redundant

The analysis show that the income group of households determined their likelihood of choosing between detached and compound housing. Overall, households with relatively lower incomes were more likely to live in compound than in detached house. For example, with an odds ratio of 0.140, low-income households were 81% less likely to live in detached house than in compound housing. As the odds ratio for lower-middle (0.192) and upper-middle (0.165) income households show, the likelihood of households living in compound housing diminishes as income increases. Moreover, households with heads having low levels of educational attainment, with odds ratio of 0.164, were 84% less likely to live in a detached house than in a compound house. Living in the historical-core of the metropolis also decreased the likelihood of households obtaining accommodation in detached housing.

Between semi-detached and compound housing, the MLR analysis show that that households belonging to the low-income, lower-middle-income and upper middle income groups with odd ratios of 0.117, 0.147 and 0.198 respectively, were less likely to find accommodation in a semi-detached house than in a compound house. Controlling for income, however, households with married heads, had odds of living in a semi-detached house, almost three-times higher than living in a compound house. Furthermore, households which have heads, who have low and moderate levels of education with odds ratios of 0.321 and 0.342 respectively, were less likely live in a semi-detached house than in a compound house controlling for the other factors.

The final pair of comparison of households' dwellings is between living in flats in multi-storey buildings and living in a compound house. Again, low-income, lower-middle-income and upper-middle-income households with odd ratios of 0.130, 0.085 and 0.176 respectively, were more likely to reside in compound housing than in flats in multi-storey buildings. Moreover, controlling for income, households with heads having low and moderate levels of educational attainment with odds ratio of 0.380 and 0.511 were 62% and 48.9% less likely to live in a flat than in a compound house respectively.

In summary, the analysis has shown a significant relationship between attributes of the households and the types of dwellings they choose to live in. Household attributes such as income, marital status and levels of educational attainment of heads, had statistically significant effect on the types of dwelling they found accommodation in. Overall, as income and educational attainment of household heads increase, they tend to prefer living in detached and semi-detached housing compared to living in compound housing and flats. Controlling for income, couples preferred semi-detached houses to other types of housing in the metropolis. Households of lower and middle level incomes, living in the historical-core neighbourhoods of the metropolis, on the other hand, tend to choose traditional compound houses.

4.7.2 Examining housing tenancy choice: a multinomial logistics regression

In this section, the determinants of housing tenancy choice are examined using MLR analysis. A summary of the variables used in the analysis is presented in table 4.11. Specifying home-ownership as the reference category, the analysis examines the likelihood of a household renting or living rent-free as compared to owning their accommodation.

Table 4.11: Variables in logistics regression analysis of dwelling types preferences

	Variable name	**Type and coding**
Dependent variable	Housing tenancy	Categorical coded: 1- owner-occupied; 2- renting; 3- rent-free
Independent variables	Income-group	Categorical coded: 1-low; 2- lower-middle; 3-upper-middle; 4-high; and 5-Rich
	Marital status	Categorical coded: 1-single; 2-married; 3-co-habiting; and 4-single (divorced or widowed)
	Family size	Scale variable
	Educational attainment of household heads	Categorical coded: 1-Basic School; 2- Secondary & Post-Secondary; and 3- Tertiary
	Residential zone of residence	Categorical coded: 1-Historical-Core; 2- Inner-suburb; and 3- Outer-suburb

The likelihood ratio estimates from the MLR analysis of tenancy choice is summarized in table 4.12.

Table 4.12: Logistics regression analysis of the likelihood of tenancy choice by households

		95% Confidence Interval for Exp(B)		
Renting vs. Owner-occupier	*b(SE)*	Lower Bound	Odds Ratio	Upper Bound
Intercept	2.545 (.879) *			
Low-income	2.276(.762) *	2.188	9.735	43.322
Lower-middle	1.149(.639)	0.901	3.154	11.040
Upper-middle	1.207(.617) *	0.997	3.342	11.200
High	.842(.609)	0.703	2.321	7.662
Rich	0^b			
Detached	-3.414(.412) ***	0.015	0.033	0.074
Semi-detached	-1.493(.445) **	0.094	0.225	0.537
Flat	-.897(.507)	0.151	0.408	1.102
Compound	0^b			
Single (never married)	.125(.587)	0.359	1.133	3.581
Couple (married)	.116(.423)	0.490	1.123	2.573
Couple (co-habiting)	-.034(.860)	0.179	0.966	5.216
Single (divorced and widowed)	0^b			
Basic School	-.499(.347)	0.307	.607	1.199
Secondary & Post-Secondary	-.384(.350)	0.343	.681	1.354
Tertiary	0^b			
Family size	-.267(.087) *	0.646	0.765	0.907
Historical-Core	1.182(.421) *	1.427	3.259	7.443
Inner-suburb	.189(.295)	.677	1.209	2.157
Outer-suburb	0^b			
Rent-free vs. Owner-occupier				
Intercept	2.950 (0.871) *			
Low-income	1.734(0.745) *	1.317	5.666	24.382
Lower-middle	.549(0.618)	0.515	1.731	5.814
Upper-middle	.175(0.604)	0.365	1.192	3.893
High	.147(0.591)	0.364	1.159	3.691
Rich	0^b			
Detached	-3.033(0.413) ***	0.021	0.048	0.108
Semi-detached	-1.943(0.472) ***	0.057	0.143	.361
Flat	-1.603(0.535) *	0.071	0.201	0.574
Compound	0^b			
Single (never married)	-.545(0.594)	0.181	0.580	1.858
Couple (married)	-.574(0.415)	0.250	0.563	1.272
Couple (co-habiting)	-.667(0.873)	0.093	0.513	2.840
Single (divorced and widowed)	0^b			
Basic School	-.148(0.365)	0.422	0.862	1.762
Secondary & Post-Secondary	.092(0.364)	0.537	1.096	2.238
Tertiary	0^b			
Family size	-.186(0.090) *	0.696	0.831	0.992
Historical-Core	1.133(0.431) *	1.336	3.106	7.223
Inner-suburb	.142(0.308)	0.631	1.153	2.108
Outer-suburb	0^b			

Note: R^2 = 0.286. (Cox and Shell), 0.329 (Nagelkerke), *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$,
[b] This parameter is set to zero because it is redundant

Firstly, the model fitting information indicated the final model is a better fit than the original model (X^2 = 216.499, df = 32, $p < 0.001$) while the Goodness-of-fit measures showed that the final model was a good fit to the data (Pearson statistics: X^2 = 894.891, df = 900, p = 0.542; and Deviance statistics: X^2 = 823.175, df = 900, p = 0.968).

Comparing the likelihood of home-ownership with renting shows that low-income and upper-middle-income households had odds of renting 9.74 times and 3.34 times higher than owning their homes, respectively. With odd ratios of 0.033 and 0.23 associated with residence in detached and semi-detached dwellings respectively, the result indicates that there is a small likelihood that household occupying these dwellings would rent controlling for other factors. In other words, households living in detached and semi-detached houses were more likely to own their accommodation. Moreover, living in the historical-core was associated with odds of renting 3.3 times higher than owner-occupied tenancy, controlling for income and dwelling types.

Furthermore, low-income households have odds of living rent-free, 5.69 times higher than owning their homes controlling for other factors. Other factors controlled, households were more likely to live rent-free in compound housing than in detached, semi-detached and flat housing. Indeed, households living in detached (odds ratio = 0.048), semi-detached (odds ratio = 0.143) and flats (odds ratio = 0.201) were 95%, 86% and 80% respectively, less likely to live rent-free than to own their accommodation. The analysis of tenure choice further show that households living the historical-core neighbourhoods of the metropolis had odds of living rent-free, 3 times higher than owning their accommodation. Finally, controlling for other factors, the odds of living rent-free (odds ratio = 0.83) decreases as family size increases.

In summary, the MLR analysis has shown that household income, dwelling type occupied, urban-zone of residence and family size have statistically significant effect on housing tenancy choice in the Kumasi metropolis. Overall, lower and middle income households, were more likely to rent their accommodation or live rent-free. Also, renting and rent-free tenure arrangements were common in compound housing within the historical-core of the metropolis. On the contrary, households living in detached and semi-detached housing, particularly suburban neighbourhoods of the metropolis were more likely to own their accommodation than rent or live rent-free. Marital status and educational attainment of household heads, however, did not have statistically significant effect on the tenancy choice.

4.7.3 Occupancy characteristics

Besides the importance households attach to location defining attributes as well as preferences for dwelling types and housing tenancy, occupancy characteristics realized at their place of residence could shed further light on the residential location choice process. In this regard, the analysis examined the room occupancies within the different housing types and among the different households in the metropolis (see table 4.13). The analysis found positive association between family size and the number of rooms occupied ($r = 0.276$, $P < 0.001$), income group and rooms occupied ($r = 0.401$, $P < 0.001$) and a negative association between dwelling types[10] occupied and number of rooms ($r = -0.434$, $P < 0.001$).

Table 4.13: Number of rooms occupied by households.

		Mean	Standard Deviation
Dwelling Type	Detached	4	2.201
	Semi-detached	2	2.068
	Flat	2	2.917
	Compound	1	1.67
	All	2	2.217
		Mean	**Standard Deviation**
Income Group	Urban poor + Low-income	1	1.558
	Lower-middle-income	2	1.909
	Upper-middle income	2	1.854
	High income	3	2.252
	Rich	5	3.494

Source: Based on Field Survey, February 2015

A linear regression model was fitted to the data to ascertain the effects of income, family size, marital status and dwelling types on room occupancy **(**see table 4.14)[11].

Table 4.14: Determinants of rooms occupancy in the Kumasi metropolis[12]

Predictors	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	1.995	.418		4.768	.000
Income	.325	.071	.182	4.558	.000
Family size	.234	.046	.197	5.085	.000
Coupling status	.059	.085	.024	.693	.489
Dwelling type	-1.113	.155	-.263	-7.190	.000

Note: $R^2 = 0.211$, $p<0.001$.

[10] Dwelling type was dichotomized as either compound or non-compound. This means that rooms available to households' decreases as occupancy shifts from detached, semi-detached and flat to being compound housing.
[11] See appendix 1 for scatter plots of the relationship between the variables in the model. The scatter plots show that the data is rather noisy, as such, the relationships among the variables are only weakly expressed in the regression analysis.
[12] The explanatory variables 'coupling status' and 'dwelling-type' are dichotomous. Dummies or indicator values (i.e. 1, 0) were created to assign the numerical values to levels of categorical variables and to avoid errors resulting original numerical codes being interpreted as numerical values in the regression.

The results of the analysis show that the number of rooms occupied by households increases by 0.325 as income category increases from low to rich. As family size increased by one, the number of rooms occupied by households increased by 0.234, controlling for all other factors. On the contrary, the number of bedrooms occupied by households in compound housing decreases by one controlling for other factors. Marital status, however, had no statistically significant effect on the number of rooms occupied by households. The regression model explains 21% of the variability in the number of rooms occupied by households.

4.7.4 Rental levels and rent-to-income ratio

Property price or rents constitute another dimension of housing choice. Given that home-ownership in the case study context is achieved through self-build incremental building after land acquisition, data on house prices is not readily available. Even if such that existed, it would not make any meaningful contribution to the analysis of housing choice for reasons stated above. On the contrary, data about rental market characteristics could be elicited directly from the households. Using the reported house rent data, rents were divided into bands based on percentile distributions (i.e. from the 1st to the 100th percentile rent). The analysis yielded 10 different rent-bands (see table 4.15).

Table 4.15. Rent Bands in the Kumasi Metropolis.

Rent Bands	Description	
	Percentile	Monthly Rent (GH₵)
Band-A	1st - 8th	5-18
Band-B	15th - 25th	20 - 25
Band-C	38th - 48th	30- 35
Band-D	56th - 73rd	40-60
Band-E	77th - 82nd	70-90
Band-F	85th - 92nd	100 -150
Band-G	94th - 96th	160 -200
Band-H	97th - 98th	250 - 300
Band-I	98th - 99th	400 - 500
Band-J	100th	600 -700

Source: Based on Field Survey, February 2015

To ascertain the determinants of households rent levels in their current accommodation, a linear regression was fitted to the data. The explanatory variables used in the model were number of bedrooms, tenancy contract period, dwelling type (i.e. detached, semi-detached, flat and compound); residential class[13] (high, medium and low class residential areas); and income of households. The result is summarized in table 4.16.

Table 4.16: Determinants of house rents in the Kumasi Metropolis

	Unstandardized Coefficients		Standardized Coefficients		
Predictors	B	Std. Error	Beta	**t**	**Sig.**
(Constant)	118.449	30.592		3.872	.000
Number of bedrooms	11.486	4.227	.159	2.717	.007
Tenancy period (years)	-5.526	1.617	-.191	-3.417	.001
Dwelling type	-13.465	4.530	-.170	-2.972	.003
Residential class	-19.712	8.219	-.135	-2.398	.017
Income group	11.892	4.320	.159	2.753	.006

Note: $R^2 = 0.161$, $p < 0.001$

The results show that for everyone unit increase in the number of bedrooms occupied by households, monthly rent increased by GH₵11.486 controlling for other factors. Longer tenancy periods were associated with GH₵ 5.526 decrease in monthly rents controlling for other factors. In addition, change from detached towards compound was associated with GH₵ 13.465 decrease in monthly rent. This means that rent levels are expected to be lower in compound houses compared to detached and semi-detached, and flats. It is also because for most households living in compound houses, they occupy a single room. The analysis further show that as residential class changes from high-class to low class neighbourhoods, monthly rent decreases by GH₵ 19.712. Finally, as household income increases from low income to rich, house rent increases by GH₵ 11.892, meaning that households of higher income tend to pay more for rent.

Moreover, as part of the survey, households in the rental sector were asked to indicate how much they paid as rent per month and whether they considered the rent paid affordable. Most households (95%) indicated that their rents were affordable to them relative to their incomes. Indeed, Spearman's Rho correlation analysis found a moderate positive association between income-group of households and rent band of their accommodation ($r = 0.362$, $p < 0.001$). Using this information, the income-to-rent-ratio of each household group was computed (see table 4.16). In calculating the income-rent-ratio, the earnings of households under each of the income groups were divided into two ranges, lower half and upper half as shown in the second and third columns of table 4.17. Next, the rent band for each household within the two

[13] The definition of residential class as high, medium and low was determined by the income groups, dwelling types and prestige of the 32 neighbourhoods from the households were interviewed.

categories was determined based on the monthly rent paid (4th column). For each income group, the percentage of households paying any given rent-band was identified (5th column). The corresponding rent-to-income-ratio, expressed in percentage terms was computed for each household in each income category.

Table 4.17: Rent to income ratios of households living in rented housing

Income Group	Income range GH₵ (Lower half)	income range GH₵ (upper half)	Rent-Band	% of households in rent band	% Rent-Income ratio (Lower half Income)	% Income Rent-Income (Upper half Income)
Low income	150-350	400-700	A-B	46	3 - 7	1 - 4
			C-D	43	17 - 20	8 - 9
			E-F	11	43 -47	18 - 21
Lower-middle	750-1050	1100-1200	A-B	30	0.7 - 2.4	0.5 -2.1
			C-D	59	4 - 2.4	2.7 -5.0
			E-F	11	9.3 - 14.3	6.4 - 12.5
Upper-Middle	1250 -1660	1680-2000	A-B	23	0.4 - 1.5	0.3 -1.3
			C-D	39	2.4 - 3.6	1.8 -3.0
			E-F	33	5.6 - 9.0	4.2 -7.5
			G-H	4	12.8 - 18.	9.5 -17.9
			F	1	30 -32	
High	2050 -2700	2720-4000	A-B	12	0.2 - 0.9	0.2 - 0.6
			C-D	39	1.5 - 2.2	1.1 - 1.5
			E-F	27	3.4 - 5.6	2.2 - 3.8
			G-H	18	7.8 - 11.1	5.9 - 7.5
			F	4	14.7 - 18.5	
Rich	4100- 4700	5600-8000	D	43	1 - 1.3	0.7 - 0.8
			F	57	2.4 - 3.2	1.8 - 1.9

Source: Based on Field Survey, February 2015

The results of the rent-income ratio analysis showed that within each broad income group identified, the rent-bands of accommodation and hence the percentage of income households spent on housing varied. For example, among low-income households, the majority (46%) spent GH₵ 5 to GH₵ 25 (i.e. rent bands A-B) every month on house rent. This represented between three percent and seven percent of the monthly incomes low-income households earning the lower-half income range—GH₵ 150 to GH₵ 350 respectively. For low income households paying rents band A-B and earning the upper half income range (i.e., GH₵ 400-GH₵ 700), their rent-income ratio was one percent and four percent respectively. The rent-income ratio for lower middle, upper middle, high income and rich households were also computed.

Besides contributing to understanding the full range of choices households make as part of their residential location choice, the rental market information presented in this section will be applied later in Chapter seven to simulate the location choice process.

4.8 Job market characteristics and employment location choice

This section addresses second pair of the urban location choice process, which is the job location of working members within the households interviewed. The analysis is based on data obtained from the 1,158 workers interviewed from the household sample of 665. A brief discussion of the job market and employment situation in the metropolis based on the survey data is provided followed with analysis of the distribution of job locations and the determinants of workers' employment locations.

4.8.1 Employment location, work industry and skills of workers

The job locations of workers were first categorized into two namely; home-based and non-home-based. The former refers to all employment located within the home and or located not more than 100m (0.1km) within the immediate vicinity of the home while the latter refers to all employment located outside the home in one of the major employment zones within the metropolis. Out of the 1,158 workers interviewed, 331, representing 30% had home-based employment while the remaining 827 (70%) had non-home-based employment.

The work industry as well as the skills levels of workers were also examined (see figure 4.5). A disproportionately larger share of all employment (88%) was in the service sector. Activities in the service sector were mainly commercial ventures involving small to medium scale retail trade in the informal economy, and formal sector jobs involved in rendering non-tangible services in the administration, education, health, banking and finance sectors.

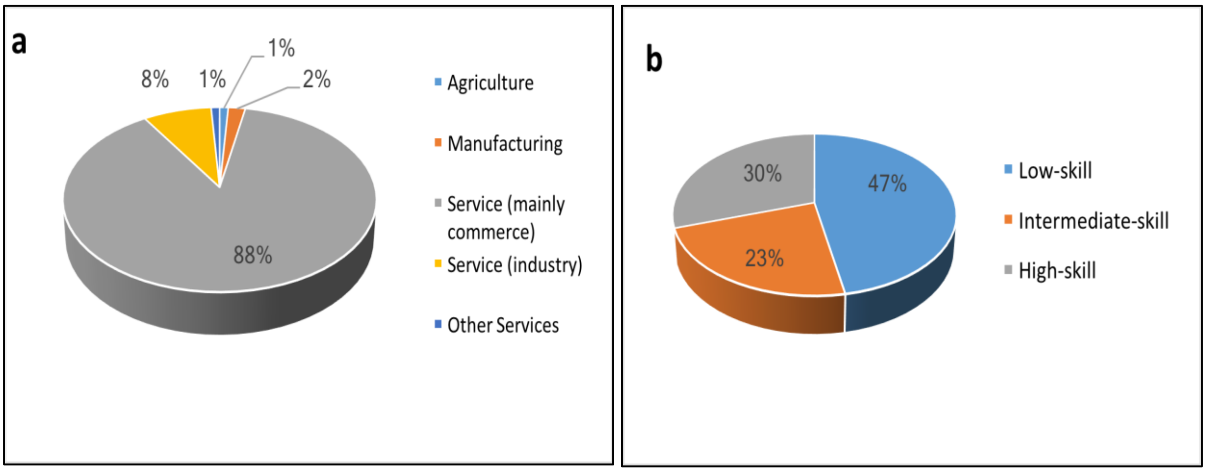


Figure 4.5: (a) work industry and (b) skill levels of workers. Source: Based on Field Survey, February 2015

The analysis also distinguished two others type of service sector employment, which were those involved in rendering tangible and non-tangible services high up the value chain, and those involved in military and security services. The former includes specific activities such as waste management services, construction, utilities, transportation and warehousing, and represented eight percent of employments among the workers interviewed. Manufacturing and agricultural jobs constituted only two percent and one percent of employment among the respondents.

Furthermore, nearly half of all jobs (47%) were low-skilled activities mainly in the informal sector. These kinds of jobs require no formal training or apprenticeship. Inter-mediate level skill jobs, which require formal qualifications at the secondary and post-secondary levels or through the completion of apprenticeship training constituted 23% of employment among the workers interviewed. Finally, 30% of all employment were in high-skilled professional and managerial jobs such as banking, teaching, engineering, technicians and the associated professions.

Chi-square analysis found a positive association between job location and work industry of workers (see table 4.18) as well as jobs location and skills of workers (see table 4.19). Most home-based jobs (95%) and non-home-based jobs (86%) in the metropolis were commercial activities in the service sector. Moreover, in terms of home-based jobs in the service sector, the majority (68%) were low-skilled compared with 39% among non-home-based jobs (see table 4.19). Only nine percent of home-based jobs were highly-skilled. This included consultants, school owners and health practitioners who operated in offices from their homes.

Table 4.18: Relationship between jobs location and work industry

	Home-based		**Non-home-based**	
Work industry	Number	Percentage	Number	Percentage
Agriculture	2	1	7	1
Manufacturing and Mining	0	0	8	1
Service (mainly commerce)	313	95	712	86
Service (industry)	13	4	88	11
Other Services	3	1	12	1
Total	**331**	**100**	**827**	**100**
X^2= 22.753, df = 4, $p < 0.001$				
Cramer's V = 0.140, $p < 0.001$				

Source: Based on Field Survey, February 2015

Table 4.19: Relationship between jobs location and skills of workers

Job Skill	All jobs		Home-based		Non-Home-based	
	Number	percent	Number	percent	Number	percent
Low-skill	544	47	224	68	319	39
Intermediate-skill	262	23	79	24	183	22
High-skill	353	30	28	8	325	39
Total	**1158**	**100**	**331**	**100**	**827**	**100**
$X^2 = 83.719$, df = 2, $p < 0.001$						
Cramer's V = 0.270, $p < 0.001$						

Source: Based on Field Survey, February 2015

4.8.2 Spatial distribution of home-based and non-home based jobs

Firstly, with regards to home-based employment, the analysis found that whereas 40% were in the historical-core, the rest were distributed equally between the inner-suburban and outer-suburban-zones of the metropolis. Non-home based jobs were distributed among the five major employment zones and other sub-employment centres in the metropolis. The location of these employment zones is shown in figure 4.6.

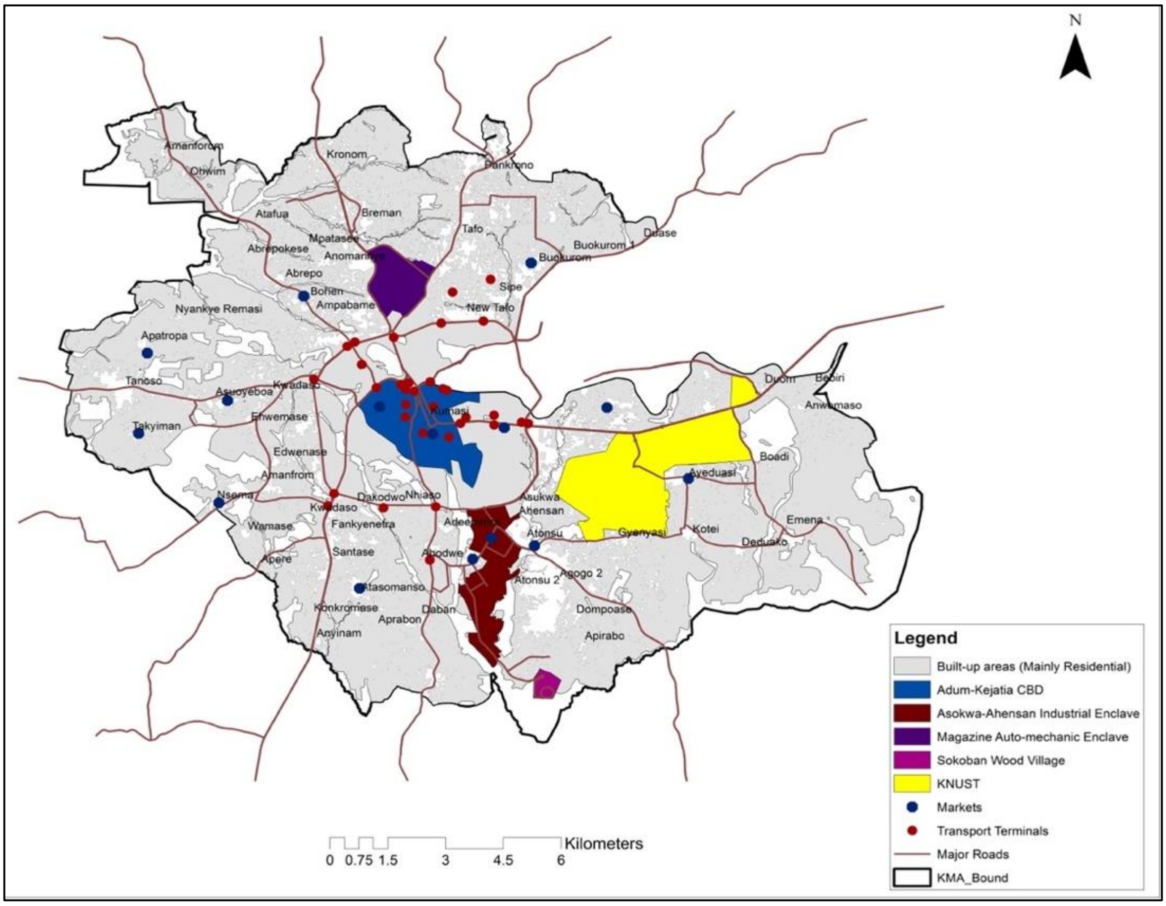


Figure 4.6: Location of Employment Zones.
Source: Based on data obtained from the TCPD, KMA

Workers in the manufacturing sector had their job locations in one of the two industrial enclaves—Asokwa-Ahensan Industrial Enclave and the Sokoban Wood Village. The distribution of non-home-based jobs in the service sector among the major employment zones is summarized in table 4.20. Non-home based jobs in the service sector, were however widely distributed across the major employment zones in the metropolis. Overall, 94% of non-home-based service jobs of workers were located within the Kumasi metropolis while the remaining six percent were located outside the metropolis.

Table 4.20: Distribution of non-home-based service jobs by employment zones

Employment Location	Number	Percent
Employment zone-1 (Adum/Kejetia CBD)	393	48
Employment zone-2 (Asokwa-Ahensan Industrial Enclave)	28	3
Employment zone-3 (Magazine auto-mechanic enclave)	206	25
Employment zone-4 (Sokaban Wood Village)	20	2
Employment zone-5 (KNUST, University)	60	7
Around markets and terminals	76	9
Outside KMA	29	6
Total	**812**	**100**

Source: Based on Field Survey, February 2015

Among those located within the metropolis, nearly half (48%) were in the Adum-Kejetia CBD area followed by 25% in the Magazine auto-mechanic enclave. Moreover, about three percent and two percent of non-home-based service jobs were located within the Asokwa-Ahensan Industrial Enclave and Sokoban Wood Village, respectively. Other major employment locations included the KNUST, which accounted for seven percent of the non-home-based service job locations of workers. Also, about nine percent of non-home-based jobs in the service sector clustered around activity nodes such as local markets and transport terminals located across the metropolis.

4.8.3 Determinants of job location choice

A binary logistics regression analysis was conducted to ascertain the determinants of employment location. Table 4.21 provides a summary of the variables in the regression analysis. The outcome variable was home-based and non-home-based job locations in the service sector. Agricultural and manufacturing jobs were excluded from this analysis because as indicated earlier, these had non-home-based locations. Thus, the total sample for this analysis was the 1, 126 individual service workers interviewed.

Table 4.21: Variables in the binary logistic regression of job location

	Variable name	Type and Coding
Dependent variable	Job location	Categorical: 1-home-based; 0-non-home-based
Independent variables	Income-group	Categorical: 1-Low (below 25th percentile); 2-middle (percentile and 75th percentile) and 3- high (above 75th percentile)
	Skill	Categorical: 1-Low; 2-intermediate; 3-high
	Work Industry	Categorical: 1-Service (mainly commerce); 0-Other services
	Urban zone	Categorical: 1-Historical-core; 2-Inner-suburb; 3-Outer-suburb

The results of the analysis **(**see table 4.22) show that workers engaged in commercial activities in the service sector have odds of their work being home-based, 2.93 times higher than those engaged in service-industry activities. Controlling for other factors, low-income workers in the service sector had odds of their work being home-based, 2.15 times higher than high-income earners in same sector. Middle -income workers in the service sector on the other hand, had odds of their work being home-based, 1.12 times higher than high-income earners in same sector.

Table 4.22: Parameter estimates of the logistic regression analysis of job location choice

		95% C.I. for EXP(B)		
Predictors	**b (SE)**	**Lower**	**Odds Ratio**	**Upper**
Work industry	1.076 (0.254) ***	1.784	2.933	4.822
High Income	0^b			
Low Income	.725(0.222) **	1.336	2.065	3.194
Middle Income	.079(0.191)	.745	1.082	1.573
High Skill	0^b			
Low Skill	1.571(0.217) ***	3.146	4.813	7.362
Moderate Skill	1.438(0.227) ***	2.701	4.214	6.575
Outer-suburb	0^b			
Historical-core	.325(0.189) *	.957	1.385	2.004
Inner-suburb	-.296(0.184)	.519	.744	1.067
Constant	-3.341(0.341) ***		.035	

Note: R^2 = 0. 182 (Nagelkerke), **p < 0.01, ***p < 0.001, *p < 0.05
[b]This parameter is set to zero because it is redundant

Moreover, the odds of low-skilled and moderate-skilled workers having their work located at home or within the immediate vicinity of the home was 4.81 and 4.21 times respectively, higher than highly-skilled workers. Overall, the odds of being a home-based worker decreases in the sub-urban zones of the metropolis. Workers living in the historical-core of the metropolis have odds of their work being home-based 1.38 times higher than workers having residence in the outer-suburban zone.

In addition to the above determinants of job location choice, individual workers indicated the importance they attached to a set of 8 evaluation items regarding their job location choice on a five point Likert scale ranging from "*very important"* to *"not-important-at-all*. Descriptive statistics of the items is presented in table 4.23.

Table 4.23: Descriptive Analysis of 8 Evaluation Items of job location choice

Evaluation Items	**Mean**	**SD**	**N**
Proximity to home/residence	2.17	1.30	1158
Opportunities for high paid jobs	2.28	1.16	1158
Availability of preferred job	2.02	1.03	1158
Availability market for my goods/services	1.91	1.05	1158
Proximity to work place of my spouse/partner	3.32	1.23	1158
Proximity to school of child/children	3.07	1.28	1158
Proximity to essential services (banking, postal etc.)	2.54	1.26	1158
Proximity to markets/shopping centres	2.30	1.22	1158

Source: Based on Field Survey, February 2015

A principal axis factor analysis was conducted on eight items to distil the eight items into components. Firstly, The Kaiser-Meyer-Olkin Measure verified the sampling adequacy for the analysis (KMO = 0.68, above the acceptable limit of 0.5). Bartlett's test of sphericity was also significant ($X^2 = 2613.314$, df = 28, $p < 0.001$). Also, the determinant of the correlation matrix was 0.104, which was bigger than the cut-off value of 0.00001, indicating that there was no problem of multi-collinearity among the variables. Initial analysis was run to obtain eigenvalues for each evaluation item in the data. Two factors had eigenvalues over Kaiser's criterion of one and in combination, explained 55% of the total variance (see table 4.24). Using the scree-plot and Kaiser's criterion for eigenvalues, as well as given the large sample size (i.e. N= 1158) the two factors were retained.

Table 4.24: Initial Eigenvalues estimates of principal factors analysis on 8 job location items

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.890	36.121	36.121	2.320	28.994	28.994	2.227
2	1.502	18.781	54.902	1.021	12.761	41.755	1.347
3	1.019	12.734	67.636				
4	.878	10.978	78.614				
5	.644	8.045	86.659				
6	.427	5.341	92.000				
7	.391	4.890	96.890				
8	.249	3.110	100.000				

Next, an oblique rotation (direct oblimin) was conducted. Table 4.25 shows the factor loadings after rotation. The results show that the items that cluster on the same factor suggest that factor

1 represents the construct, '*proximity to the home and essential amenities'* such as shopping and schools. The second factor represents *'job availability and salary levels'*. The factor correlation matrix between the two factors was 0.25. Thus, overall, individuals in their job location choice decisions gave priority to nearness of the work place to the home as well as prospects of securing work at higher remuneration levels.

Table 4.25: Factor loadings based on oblique rotation method

	Factor	
Preference Items	1	2
Proximity to home/residence	**.426**	
Opportunities for high paid jobs		**.686**
Availability of preferred job		**.773**
Availability market for my goods/services	.310	
Proximity to work place of my spouse/partner	**.577**	
Proximity to school of child/children	**.698**	
Proximity to essential services (banking, postal etc.)	**.669**	
Proximity to markets/shopping centres	**.733**	
Eigenvalues	2.890	1.502
% of Variance	36.121	18.781

Note: Absolute co-efficient values below 0.1 have been suppressed in the table
Factor loadings over 0.40 appear in bold. Source: Based on Field Survey, February 2015

4.9 Examining residential-job location choice interdependence

Empirical analysis of residential-job location independence is hinged on two major theoretical assumptions. The first posits a conditional choice process whereby individuals choose where to live first, and based on their home locations, choose where to work or vice versa. The second assumes a conjoint process whereby home and job location decisions are taken together at the same time without imposing a hierarchical structure.

The household interviews sought to gain insight into the residential-job location interdependence by verifying empirically, which of the two theoretical assumptions indicated above held true within the context of location choice decisions in the Kumasi Metropolis. This was done at two levels. Firstly, respondents were asked to indicate what their home and job location circumstances were when they first settled in the metropolis. At the second level, they were asked to reveal their last job and house moves since settling in the metropolis and to indicate the extent to which the two locations were responsive to each other during the move.

In response to the residential and job location choice relationship when they first moved into the metropolis, about 71% indicated that the decision regarding where their job locations would be in the metropolis came sometime later based on their initial places of residence. On the contrary, for the remaining 29%, who were mainly households whose heads were moving into the metropolis for the first time in response to a job offer, their residential locations decisions were taken subsequently from their respective job locations. Thus, initial location choice followed a sequential process by which the place of residence was taken first followed by decision regarding the work-place location. Indeed, the fact that the nearly half of all households are engaged in home-based employment imply that finding a suitable residential location would almost always be the most important step.

Besides the sequence of choice during their initial job locations, the respondents were also asked to reveal their last house and job moves and whether one of the locations changed in response to the other. Results from the data shown that 52% of individuals interviewed had changed residential locations within the last seven years leading up to the time the interviews were conducted. In other words, these individuals and their households had moved from a previous home location within the metropolis to their current home locations. Asked if the change of residence was the result of job change, 10% said this was the primary reason. One percent indicated that job change was a secondary reason for change of residence. For most (89%) of these respondents however, job change was not the reason at all for moving to a new house. Instead, the main triggers of residential move included a lack of comfort and privacy, particularly for those who formerly lived with extended family relations in traditional compound houses. Other reasons included tenancy contracts ending without renewal, troubles with previous landlords, the prospect of moving into the metropolis to search for new job opportunities, the need to live closer to their jobs and the need to join relatives already living in the metropolis.

With regards to recent job location changes, the results showed that whereas 15% had changed job locations in the last seven years, the remaining 85% had not done so. In 84% of the cases involving job location change, the change did not lead to residential moves. Among the few individuals who had changed job locations, only 29% indicated that the change was necessitated primarily by changes in residential locations. Whereas eight percent reported that change of residence was a secondary reason that accounted for job location change, the majority (63%) indicated that residential location change was not a reason at all for their last job location change. Instead, poor remuneration at previous jobs, job collapse leading to retrenchments and the decision to start their own private businesses were given as the major reasons for job change and hence change in job locations among most the respondents.

In summary, three key insights can be drawn from the above findings regarding the residential-job location choice interdependence. Firstly, for most households and individuals, the initial location choice process followed the sequential choice process whereby residential location decisions were taken first. With the established home locations of households as reference point, individuals make job location decisions considering proximity to the home and essential amenities, access to markets especially in the case of petty traders and sales workers in the informal economy as well as the opportunities for well paid jobs. Secondly, for nearly half (i.e. 48%), of individuals, their residential locations have remained stable since moving into the metropolis. Job locations, however, were generally more stable than residential locations given that 85% of the respondents had not changed jobs in the last seven years. The fact that job openings unlike residential vacancies are very limited and hard to find could explain the stability of job locations. Thirdly, for most of the respondents, the decision to move to a new house was not job-related, neither was the decision to change job locations triggered by residential moves in most cases.

4.10 Chapter summary

This chapter has presented an empirical analysis of the determinants of residential location choice of households, the determinants of job location choice of individual members of the household and the interdependence between the two choice sets in the Kumasi metropolis using data obtained through a cross-sectional survey.

Using principal component analysis, four main factors were identified as important macro and meso level considerations across all households in their residential location choice. These factors were; *'proximity to major infrastructure and amenities'*; '*family ties and social networks*'; '*character of neighbourhood'*; and *'proximity to core activity locations'* (i.e. work-place and school) of members of the household. Results from a series of hierarchical multinomial logistic regression models showed that factors including household income, educational attainment of heads, family size, and urban-zone of residence determined differences in dwelling type preferences and housing tenancy choice.

Furthermore, the analysis provided vital empirical insight into the job location patterns in the Kumasi metropolis. Overall, the results showed that lower income levels, residence in the historical-core of the metropolis, low-skill levels and involvement in small scale commercial activities in the informal economy were the key determinants of whether an individual had a home-based or non-home-based employment. A principal component analysis of job location factors further showed that two major factors, namely*; 'proximity to the place of residence and essential amenities'*, as well as *'job availability and higher wages'* determined individuals' job location choice.

The final section of this chapter addressed an important question regarding the residential-job location choice interdependence in the Kumasi metropolis. Results of the analysis revealed that most of the individuals interviewed considered where they would live first, and based on the residential location outcomes, decided where to work. In terms of residential and job relocation decisions, the analysis showed that generally, residential and job locations remained stable and unchanged over several years. Notwithstanding, residential relocations occurred a little more frequently than job location changes. For most of the respondents, the decision to move to a new house was not job-related, neither was the decision to change job locations triggered primarily by residential moves.

The analysis presented in this chapter will provide the backdrop for the analysis of the patterns of daily spatial interaction between the home and work-place in the case study metropolis. The next chapter deals with this subject, focusing on work trip frequency, work trip production and attraction patterns, travel mode choice and commuting times and travel costs.

CHAPTER FIVE: HOME-WORK MOBILITY PATTERNS, TRANSPORT MODE CHOICE AND TRAVEL COSTS—AN EMPIRICAL ANALYSIS

5.1 Introduction

Patterns of urban mobility reflect individuals' need to participate in various activities located within the urban area. As a concept, urban mobility encompasses many dimensions including trip purpose and frequency, trip origins and destinations, travel mode choice, route choice, travel distance, travel time and transport costs.

The purpose, origin and destination of daily trips are influenced by the long-term urban location decisions that have shaped the spatial distribution of urban functional activities such as the work-place, the place of residence, and ancillary facilities including shopping and recreation. The resulting, trip production and attraction patterns could therefore be quantified either as flows between specific points of activity locations or as flows between aggregate urban-zones containing many activities such as Traffic Analysis Zones (TAZs).

The daily human interactions with urban activity locations at the different spatial scales also involves several short-term decisions including travel mode choice, route choice and the time of day in which trips are planned to begin and end. The prevailing urban structural conditions, interact with the socio-economic characteristics, attitudes and lifestyle preferences of the individuals involved to shape these choices.

The objective of this chapter is to examine empirically, mobility patterns in the Kumasi metropolis, the case study area for this research. It examines and quantifies mobility patterns associated with the residential and job location combinations of individual workers in the metropolis. The empirical analysis of mobility patterns presented here, therefore, builds on initial analysis of residential and job location patterns in the metropolis presented in Chapter four. The analysis examines the various aspects of flows and interactions between the home and work location pairs of individuals including home-work trip frequency, trip origins and destinations, travel mode choice, travel distance, travel time and commuting costs.

5.2 Overview of the data and statistical analysis methods

The analysis of home-work mobility characteristics presented in this chapter is based on travel data obtained from a sample population of 1,558 workers through a cross-sectional survey[14] in the Kumasi Metropolis. These individuals were adult working members from the 665 randomly selected households that constituted the sample population for this research. Thus, in addition to obtaining travel data at the level of the individual, the analysis is positioned within the wider household to which the individuals are members to explore the extent to which household attributes such as income, family size, car ownership and residential locations affect individual travel choices. In addition to the primary data on travel, secondary data on Traffic Analysis Zone (TAZ) designations was obtained from the Urban Roads Department of the metropolis. The TAZ dataset, comprising aggregate zones delineated based on unique functional characteristics, constitutes the spatial scale to which the analysis of work trip production and attraction patterns obtained from the travel data is anchored.

A detailed discussion of the statistical analysis methods used in this chapter, including mathematical representations and assumptions has been presented in Chapter three, the methodology chapter covering the two empirical studies presented in this thesis. In view of this, the themes of the analysis and the methods used is presented briefly in this section.

Both descriptive statistics and regression analysis methods are employed in the data analysis. Home-work trip production and attraction among the TAZs are extracted from the travel data and analysed using basic descriptive statistics. To examine the determinants of travel mode choice, a series of Binary Logistic Regression (BLR) models are fitted to the data. The first BLR analysis examines the determinants of private car ownership among households of the individuals interviewed. The method is also applied to examine the determinants of active and motorized transport mode choice, as well as choice between different public transport modes available in the metropolis. Finally, using descriptive statistics, correlation analysis and linear regression methods, data on travel times and costs associated with home-work commuting and their determinants are analysed.

[14] Detailed discussion of the sampling approach, study variables and the survey instrument that was used to elicit data on individuals' mobility characteristics is presented in Chapter three, which is the main methodology chapter covering the empirical analysis in this chapter.

5.3 Chapter organization

The remainder of this chapter is organized into five main sections. To provide background understanding of the urban structural conditions within which daily mobility patterns in the metropolis occur, the analysis opens with a discussion of the metropolitan functional structure and the existing TAZ designations in the metropolis. Next, analysis of home and work trip production and attraction patterns at the level of the TAZs is presented. In the third section, individual workers' transport mode choice considerations are analysed followed with the analysis of the determinants of transport mode choice in the fourth section. The transport mode choice analysis focuses on the determinants private car ownership and use, active and motorised transport choice, and choice between public transport options by commuters in the metropolis. The penultimate section presents an analysis of travel times and costs associated with home-work commuting in the metropolis. The final section of this chapter provides a summary of the results presented highlighting the key findings.

5.4 Metropolitan functional structure and traffic analysis zone system

The metropolitan functional structure reflects the spatial specializations and functional linkages resulting from the distribution of major land use activities in the metropolis. It shows the existing conditions of opportunities and constraints that provide the structural framework for human spatial interaction. The spatial distribution of key metropolitan functions is depicted in figure 5.1. It shows five major functional areas can be derived from the present physical structure of the metropolis.

The main commercial and service node of the metropolis is centrally located at the Adum-Kejetia area. Functioning as the Central Business District (CBD), the Adum-kejetia area is the focal point for wholesale and retail activities, as well as administrative functions in the public service. In addition to the CBD, there are three major industrial areas in the Kumasi metropolis namely; Asokwa-Ahensan Industrial Enclave, Magazine Auto-mechanic Enclave and the Sokaban Wood Village located at the southern, north and south eastern parts of the metropolis respectively. There are also several educational establishments in the metropolis, the most important of which is the Kwame Nkrumah university of Science and Technology (KNUST), located at the eastern part. Together, these nodes constitute the major activity locations and employment areas in the metropolis.

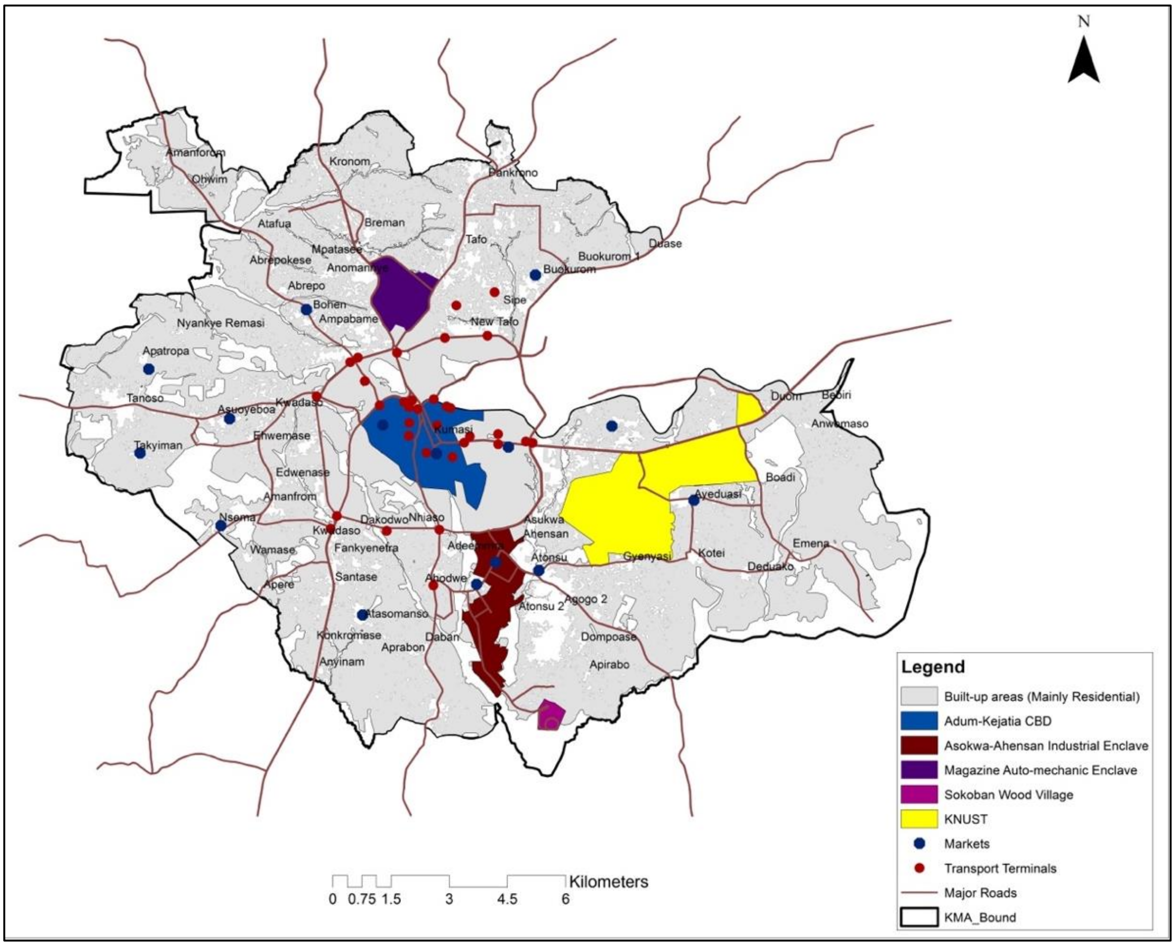


Figure 5.1: Metropolitan functional structure. Source: Based on data obtained from the TCPD, KMA

Other important functional nodes are the residential areas which depend on and support the major employment zones. The Town and Country Planning Department estimates that about 45% the metropolis's land is dedicated to residential use. Residential sectors of different characteristics are identifiable in the metropolis. This include the low-cost tenement and compound housing sector which dominate the core of the metropolis, low density and high class residential areas located to the south of the metropolis, and newly developing residential areas in the urban periphery characterized by low-density single-family detached and semi-detached buildings.

Transport passageways provide the linkages between the functional activity areas. Internally, the layout of major transport networks in the metropolis follows the traditional spoke-and-wheel format where the main avenues converge at the centre, and the inner and outer ring road system, serve as the wheel connecting the spokes (Urban Roads Department, 2004).

Furthermore, for purposes of flows and interactions analysis, the Kumasi metropolis has been divided into micro-level and macro-level TAZs. The *Urban Transport Planning and Management Studies for Kumasi* report, divides the metropolis into 29 internal micro-level TAZs (Department of Urban Road, 2004). A map showing the micro-level TAZ divisions is shown in figure 5.2[15].

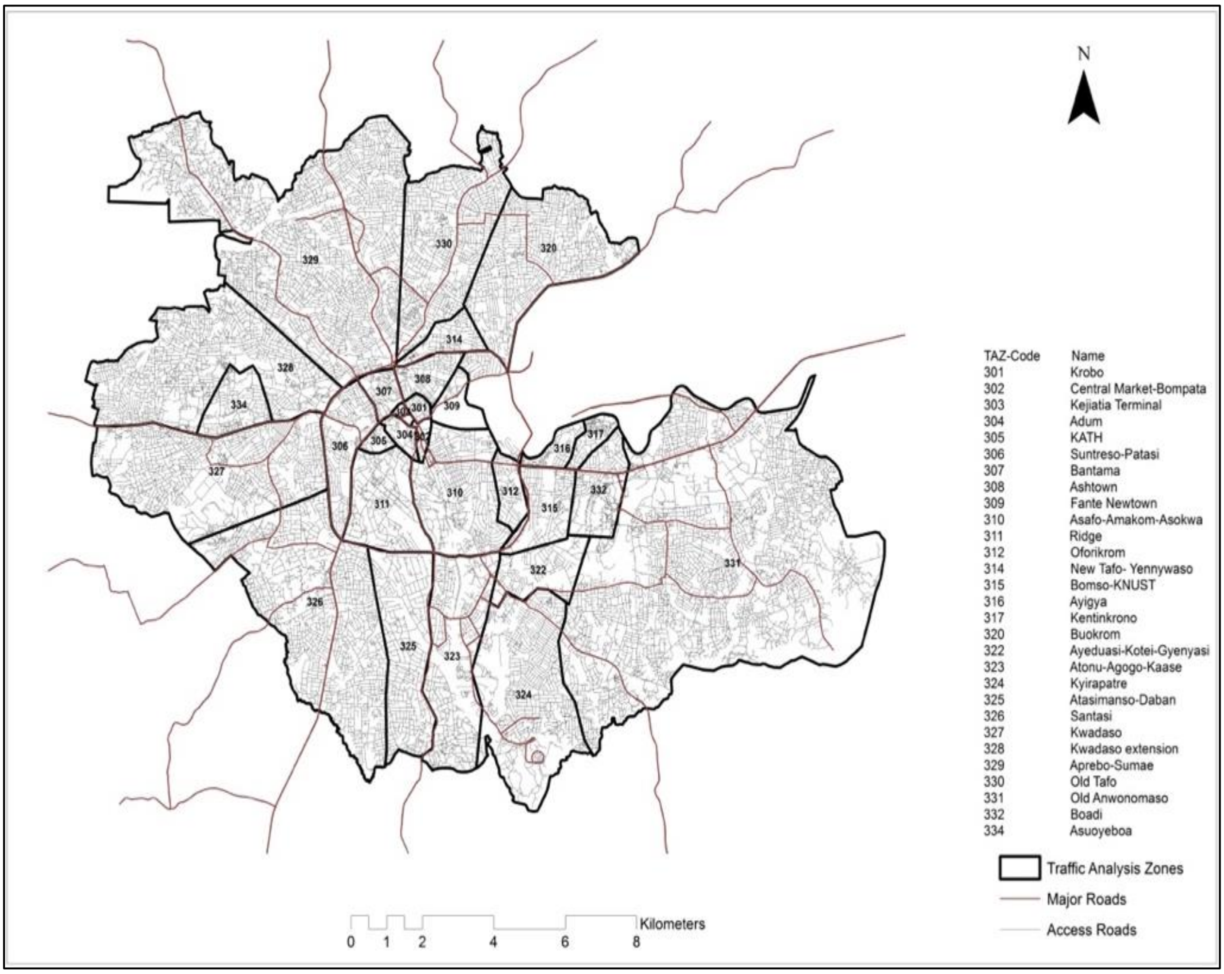


Figure 5.2: Twenty-Nine Micro Traffic Analysis Zone System in the Kumasi Metropolis

These micro TAZs are further aggregated into 6 macro TAZs as shown in figure 5.3. In the sections that follow, a brief description of each of the macro-TAZs is presented highlighting the overlapping relationships with the existing distribution of major land use functions in the metropolis presented at the beginning of this section.

---

[15] The original TAZ system comprised 32 internal TAZs. However, the Asokore-Mampong municipality has since 2010 been carved out from the KMA into a separate district. This implies that 3 of the original 32 TAZs (i.e. Aboabo-Asewase- 313, Asokore-Mampong- 318 and Duasi-321 now falls within the Asokore-Mampong district, leaving 29 Zones within the KMA.

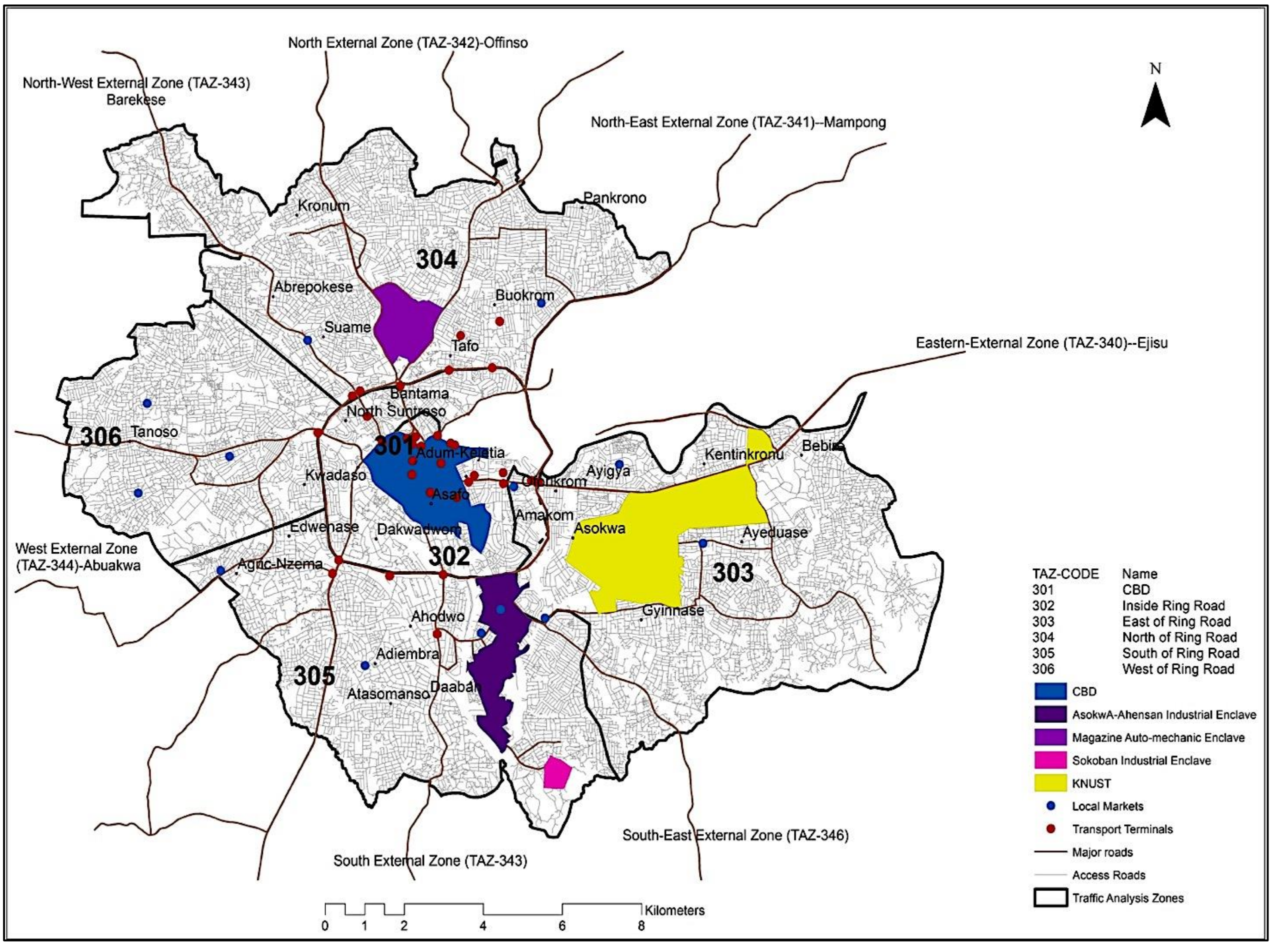


Figure 5.3: Six Macro/Aggregate Traffic Analysis Zone System in the Kumasi Metropolis

5.4.1 Central Traffic Analysis Zones—Dominantly commercial activities

As shown in figure 5.3, the centrally located zones (TAZ-301 and TAZ-302) comprises areas within the metropolitan CBD, located within the internal ring-road system where the major administrative, service and commercial functions in the metropolis are located. The dominant land use in this zone is commercial supported by mixed-residential-and-commercial buildings that house businesses as well as an estimated three percent of the total population in the metropolis (Urban Roads Department, 2004). The Kumasi Central Market, the largest open market in West Africa is also located within this zone.

The second central zone, TAZ-302, comprises areas located inside the ring road system and immediately surrounding the CBD. Located next to the CBD, this zone is characterized with mixed-land-uses of mainly commercial and residential activities in the indigenous settlements of the metropolis, including Bantama, Asafo, Ashtown and Suntreso. An estimated 24% of the total population of the metropolis live in this zone (Urban Roads Department, 2004). The

Komfo Anokye Teaching Hospital, which serves as the largest work location of workers in the health sector is also located in this zone.

5.4.2 Traffic Analysis Zones of residential activities with major industries

Overlaying the functional structure on the existing TAZ system shows that two zones (i.e. TAZ-304 and TAZ-305) have boundaries that encompasses the locations of the three major industrial enclaves and their supporting residential areas in the metropolis. TAZ-304 is located to the north of the ring-road system (see figure 5.3). In terms of industrial function, this zone hosts the Magazine Auto Mechanics Enclave which provides jobs for artisans in the car repairs and metal fabrication industry. Located in this zone are residential neighbourhoods which together accommodates an estimated 30% of the total population in the metropolis. Similarly, TAZ-305 located south of the metropolis contains industrial and residential functions. The Ahensan-Kaase Industrial enclave one of the major industrial concentrations in the metropolis is in this zone. The zone also contains medium to low density residential neighbourhoods including Atasomanso, Daaban and Santaasi located in the southern part of the metropolis. An estimated 15% of the total population in the metropolis have residence in this zone.

5.4.3 Dominantly residential Traffic Analysis Zones.

The last category of TAZs are those that contain mainly residential uses. TAZ-303, consists of medium density residential areas located east of the ring road system. The dominant land use here is residential and pockets of mixed-residential-and commercial activities located along the major arterial road that passes through it. An estimated 16% of the total population in the metropolis live in this zone. In addition to its dominant residential function, the Kwame Nkrumah University of Science and Technology, the second largest public university in Ghana is also located in this zone. Containing an estimated 12% of the total population of the metropolis, TAZ -306 comprises medium to low density residential neighbourhoods including Kwadaso, Agric Nzema and Asuoyeboa located to the west of the metropolis.

In addition to the six internal TAZs described above, the Department of Urban Roads recognizes seven external zones (see figure 5.3), which constitute external activity locations within the sub-regional context that maintain mutually beneficial functional linkages with the Kumasi metropolis. These external activity nodes include major towns such as Ejisu, Abuakwa Kodie, Esereso, Mamponteng and Asokore-mampong, located in the seven peripheral districts that share boundaries with the KMA in the sub-regional context. These neighbouring districts depend on the administrative, service and commercial functions of the KMA. In return, they serve as dormitory towns for some of the metropolis' working population, while providing jobs, mainly in the public service sector, to some of the residents of the metropolis.

5.5 Home-work trip production and attraction among traffic analysis zones

An average home-work trip frequency of six per week was recorded from the travel data (SD = 0.74, minimum = 1, maximum = 7). Data on the job location of individual workers and the home locations of their respective households was anchored to the TAZ system to derive the weekly home-work trip production and attraction patterns among the traffic analysis zones. This involved matching individually, the origin and destination TAZs associated with the weekly work trips undertaken by the respondents. Using the home TAZ as the origin, the corresponding work location TAZ constituted the destination of work trips.

Table 5.1 provides a summary of the weekly work trip origins and destinations for all trips associated with both home-based work (i.e. where the work place is the home or within the immediate vicinity of the home) and non-based work. Since home-based jobs have the same origin and destination TAZs, the origin-destinations of only non-home-based work trips is presented separately in table 5.2**.** Results of the work trip production and attraction analysis show that overall, 93% of all work trips generated in the metropolis, are distributed internally among the six TAZs. The remaining seven percent of work trips generated within the metropolis have destinations in the external TAZs.

Table 5.1: Origin and destinations of weekly trips associated with home-based and non-home-based work[16]

Origin TAZs	Destination TAZs: Internal TAZs 301	302	303	304	305	306	External TAZs 340	341	342	343	344	345	346	Total trip Origins
301	14	2	2	1	1	0	0	0	0	0	0	0	1	**21**
302	95	196	23	18	17	5	6	2	0	0	0	2	0	**364**
303	52	34	137	4	17	4	6	5	2	0	8	6	0	**275**
304	54	13	10	87	12	0	16	0	0	0	0	4	0	**196**
305	63	26	9	4	82	2	5	0	0	0	0	2	1	**194**
306	22	11	9	1	4	50	0	0	2	0	5	4	0	**108**
Total trip Destinations	**300**	**282**	**190**	**115**	**133**	**61**	**33**	**7**	**4**	**0**	**13**	**18**	**2**	**1158**

Source: Based on Field Survey, February 2015

Table 5.2: Origin and Destinations of weekly trips associated with non-home-based work only

Origin TAZs	Destination TAZs: Internal TAZs 301	302	303	304	305	306	External TAZs 340	341	342	343	344	345	346	Total trip Origins
301	8	2	1	1	1	0	0	0	0	0	0	0	2	**15**
302	63	129	21	15	10	5	4	1	0	0	0	0	0	**248**
303	45	24	102	4	12	3	4	4	1	0	5	6	0	**210**
304	42	8	8	61	8	0	12	0	0	0	0	4	0	**143**
305	37	19	4	3	56	1	3	0	0	0	0	1	1	**125**
306	17	8	9	1	4	39	0	0	1	0	3	4	0	**86**
Total trip Destinations	**212**	**190**	**145**	**85**	**91**	**48**	**23**	**5**	**2**	**0**	**8**	**15**	**3**	**1158**

Source: Based on Field Survey, February, 2015

[16] The O-D matrix, shows six origins comprising the six-internal TAZs while the destinations comprise the six-internal TAZs plus the seven-external zones. This is because the travel data was obtained from respondents in the Kumasi metropolis but not the surrounding areas.

Two main dimensions of the O-D matrix presented in Table 5.1 are examined further. The first examines the contributions, in percentage terms, of each of the TAZs to total work trip origins and total work trip destinations (figure 5.4). The analysis show that the CBD zone accounts for the smallest proportion (two percent) of total work-trip origins. In terms of trip destinations however, it attracts the largest share (26%) of all work-trips generated in the metropolis. Thus, TAZ 301, attracts work trips about 14 times higher than it generates. TAZ-302, immediately surrounding the CBD in the central locations of the metropolis accounted for 31% and 24% of total work trip origins and destinations respectively. Thus, whereas one-third of work trips originate from the two most central zones, half of all work trips generated in the metropolis terminate in these zones. This reinforces the dominant commercial and service functions of these zones.

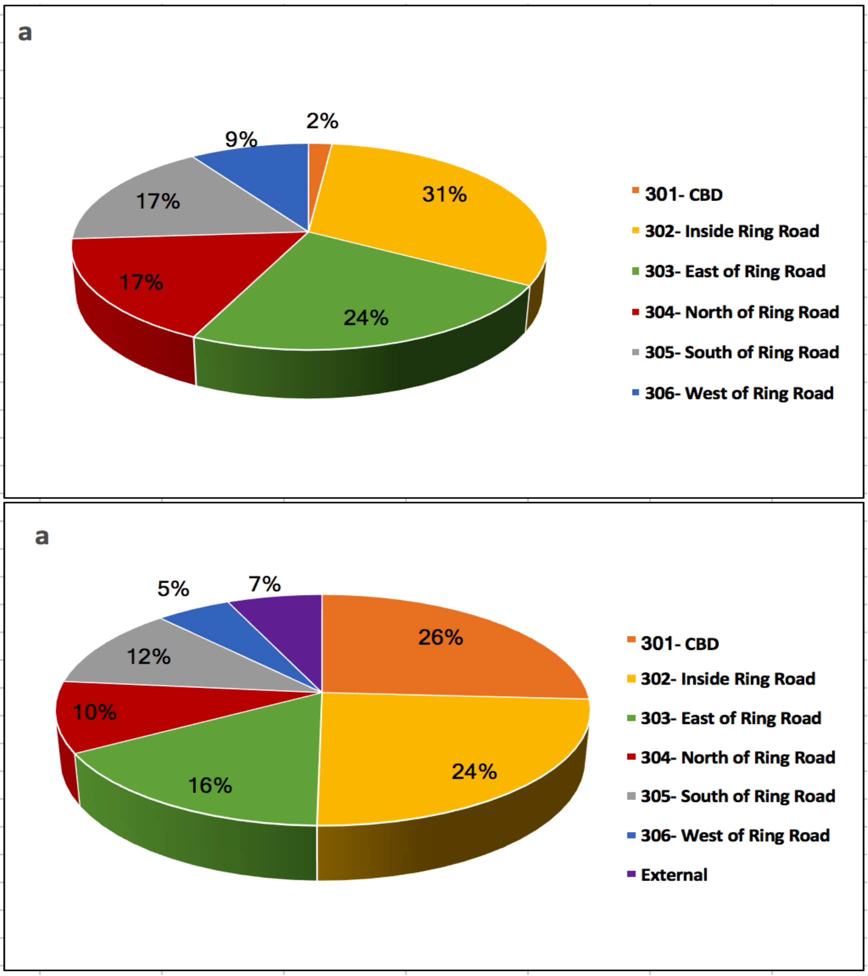


Figure 5.4: Contribution of TAZs to: (a) total work trip-origins
(b) total work trip-destinations

Moreover, the contributions of the other zones to both trip origins and destinations reflect the dual functions of these zones. For example, TAZ-303, being a dominantly residential zone accounts for nearly a quarter of total work trip origins in the metropolis. Also, being the location of a major University, this zone attracts 16% of total work trips. Given its importance as a major educational zone, the number of home-origin trips generated by this zone is about 1.4 times lower than it attracts from other locations within the metropolis. Similarly, TAZ-304 and TAZ-305 given their dual industrial and residential functions contributes each to 17% of work trip origins and 10% and 12% of trip destinations respectively. Being a major residential zone, TAZ-305 generates about 1.4 times more work trips than it attracts. TAZ-306 being a dominantly residential zone generates nearly twice more work trips than it receives.

The second dimension of the O-D matrix derived from the sample work travel data takes each of the origin TAZs and examines how the total trips it generates are distributed among all the 12 TAZs including itself. A summary of the analysis is presented in figure 5.5.

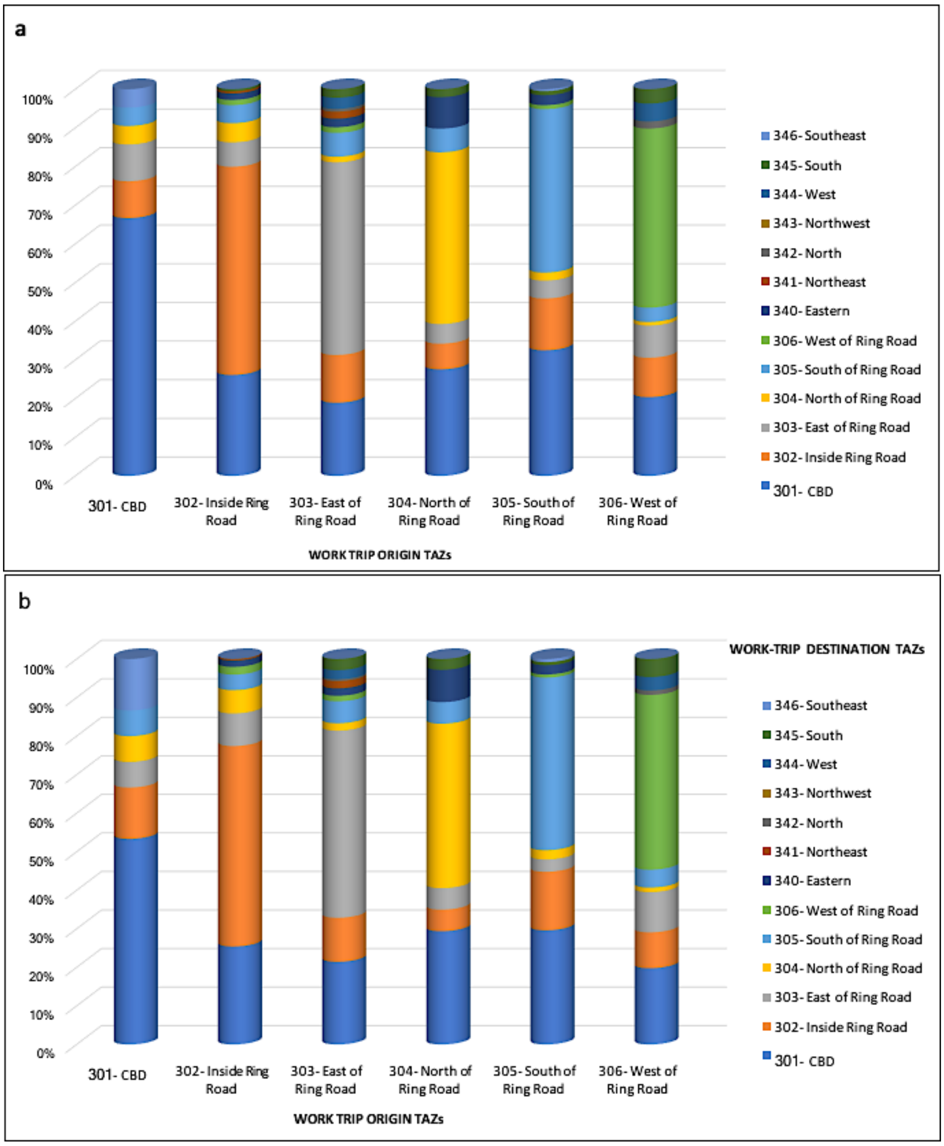


Figure 5.5: Distribution of total work trips of each origin TAZ among the destination TAZs for (a) home-based and non-home-based work-trips combined (b) non-home-based work trips only

As shown in figure 5.5, a greater percentage of all work trips have the same origin and destination TAZ. This means that, most work trips that originate from a residential location within any of the TAZs end up in a job location within the same TAZ. The trend is similar when only non-home-base jobs (figure 5.5b) are considered or when all work trips (i.e. home-based and non-home-based trips combined) are considered (figure 5.5a). For example, about 67% and 54% of all work trips starting from the two most central TAZs—TAZ301 and TAZ-302 respectively, have their destinations in these same TAZs. Similar proportions of work trip home-origins and destination patterns are observed across all the TAZs. Thus, overall, almost half of the workers interviewed in the metropolis have their home and work locations in the same TAZ.

The observed patterns where work trips begin and end in the same TAZ is partly explained by earlier findings from job location patterns presented in Chapter four. The analysis revealed that that 30% of all workers had their jobs located either in their homes or within the immediate vicinity of their dwellings at distances not exceeding 100m (0.1km). Moreover, as will be demonstrated later in the analysis of home-work travel distances, even among non-home-based workers, the home-work distance separation is relatively shorter. On the average, workers who have their jobs located in one of the major employment zones within the metropolis, lived some 4.5km from their places of work.

5.6 Travel mode preferences and choice considerations

This section focuses on the transport modes used to complete the work trips presented in the previous sections. The data on transport mode use was obtained by asking each of the 1,158 individual workers to indicate the transport mode used while going to work and returning from work in the last five days preceding the day of interview. Based on the five-day travel mode use data, the frequency of use of the different transport mode options among the respondents was established for work trips with the home as the origin (figure 5.6a) and the return legs of the trips (figure 5.6b). The respondents' frequency of use of their selected work travel modes over the five-day period have been categorized into three classes of use frequency namely; 'always'—where the mode is used every day of work in the week; 'frequent'—where the mode is used three to four times every week to work and 'seldom'—where mode is used once to twice every week to work.

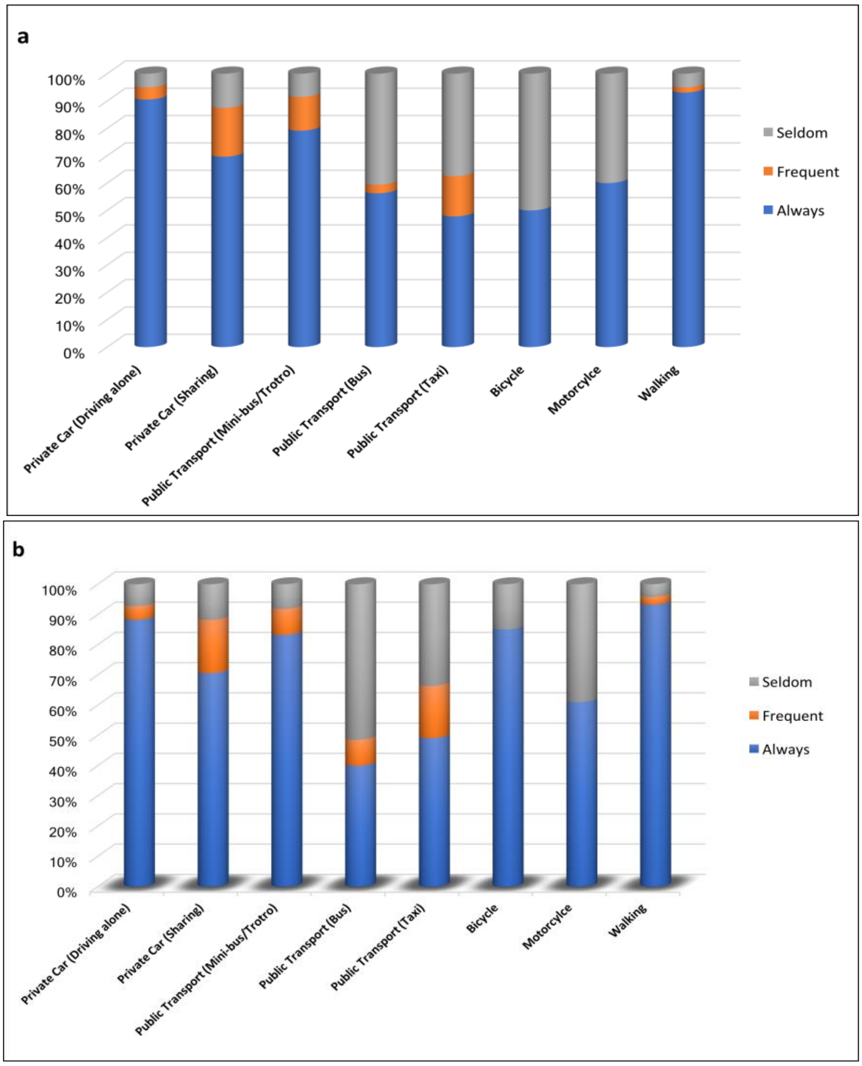


Figure 5.6: Frequency of travel mode use **(a)** from home to work **(b)** from work to home for a five-day period. Notes: **Seldom:** use mode 1-2 times every week; **frequent:** use mode 3-4 times every; and **Always:** use mode every day to work. Source: Based on Field Survey, February 2015

Overall, the data show that mode choice for both the home-origin and return leg of work journeys were consistent among the respondents across the different modes (see figure 5.6). Among private car-owners who drove alone to work, about 90% of work trips starting from the home and 88% of return trips involved driving alone in their private cars. Similarly, among private car owners who shared a ride to work with other working members of their households, about 70 % and 71% commuted this way every day for home-origin and return work trips respectively. Moreover, among public transport (mini-bus/Trotro) users about 80% of return work journey were undertaken every day of the week by this mode. In 56% and 48% of the cases, commuters using buses and taxis to work respectively used their respective modes throughout the week. This category of transport mode users, the data show, substitute between public transport in mini-bus/Trotro and taxi or buses on some days of the week. Finally, most workers (93%) who walked to work also did so consistently for all five-working days in the week.

The main work travel mode of workers was derived based on the above information. In doing so, workers using any particularly mode between three and five times a week were assigned that mode. The results, summarised in table 5.3 show that whereas 21% of workers commuted in private cars either driving alone (17%) or sharing a ride with other members of their household (4%), 45% of commuters used different modes classified as public transport. Moreover, one-third of workers chose walking as their main work travel mode.

Table 5.3: Work travel mode choice in the Kumasi metropolis

Travel Mode	**Number of commuters**	**Percentage**
Private Car (driving alone)	171	17
Private Car (sharing other household members)	49	4
Public Transport (Mini-bus/Trotro)[17]	417	36
Public Transport (Bus)	24	2
Public Transport (Taxi)	84	7
Bicycle	4	0.3
Motorcycle	6	0.5
Walking	403	33
Total	**1158**	**100**

Source: Based on Field Survey, February 2015

The survey also sought to understand the factors considered important by the commuters in choosing the above work travel mode options. To this end, individual workers evaluated the level of importance they attached to a set of factors namely; predictability of travel, affordability, comfort, privacy flexibility and travel time in relation to their preferred travel mode on a five-point Likert-scale.

As shown in figure 5. 7, travel comfort, privacy, flexibility, and quick travel times ranked very high among private car users with over 80% of them indicating these as very as very important reasons for mode choice. Less than half (49%) of private car users associated their choice with the importance they attach to affordability. Thus, the choice of the private car involves some form of trade-off between advantages it offers in terms of privacy, flexibility, comfort and travel time on the one hand and the high running cost on the other hand.

[17] In Ghana, Trotro are 10-19 seater are privately owned minibuses used as public transport.

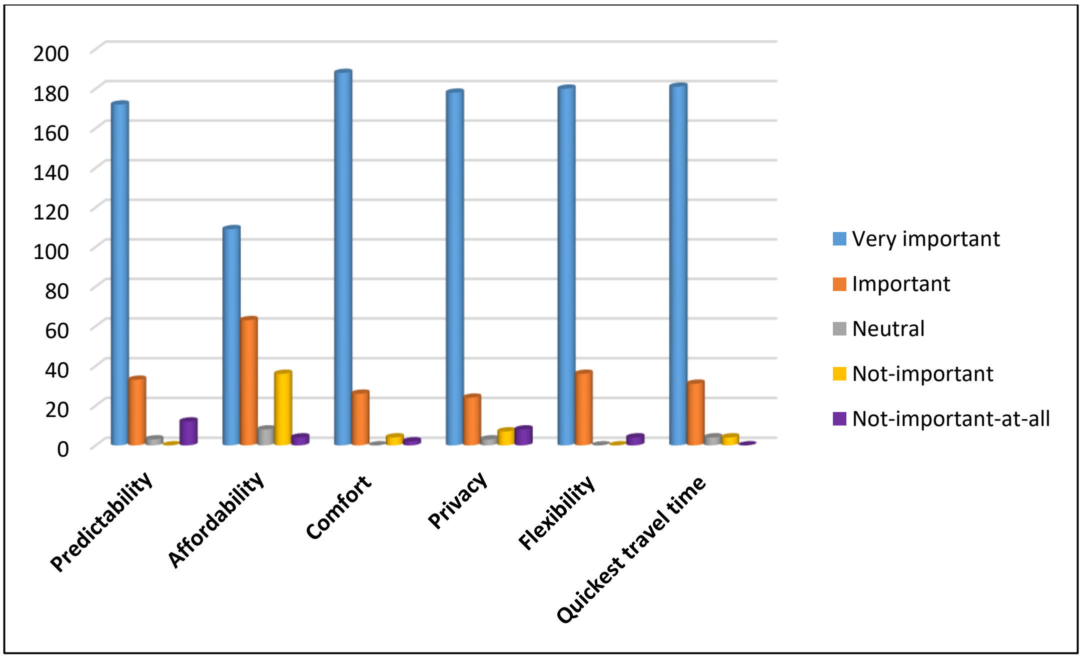


Figure 5.7: Travel mode choice considerations by private car users
Source: Based on Field Survey, February 2015

Among public transport users, 52% and 60% indicated predictability and affordability respectively, as the most important reasons for commuting to work using this mode (see figure 5.8). For the majority (84%), public transport to work was available within five minutes of getting to the bus stop.

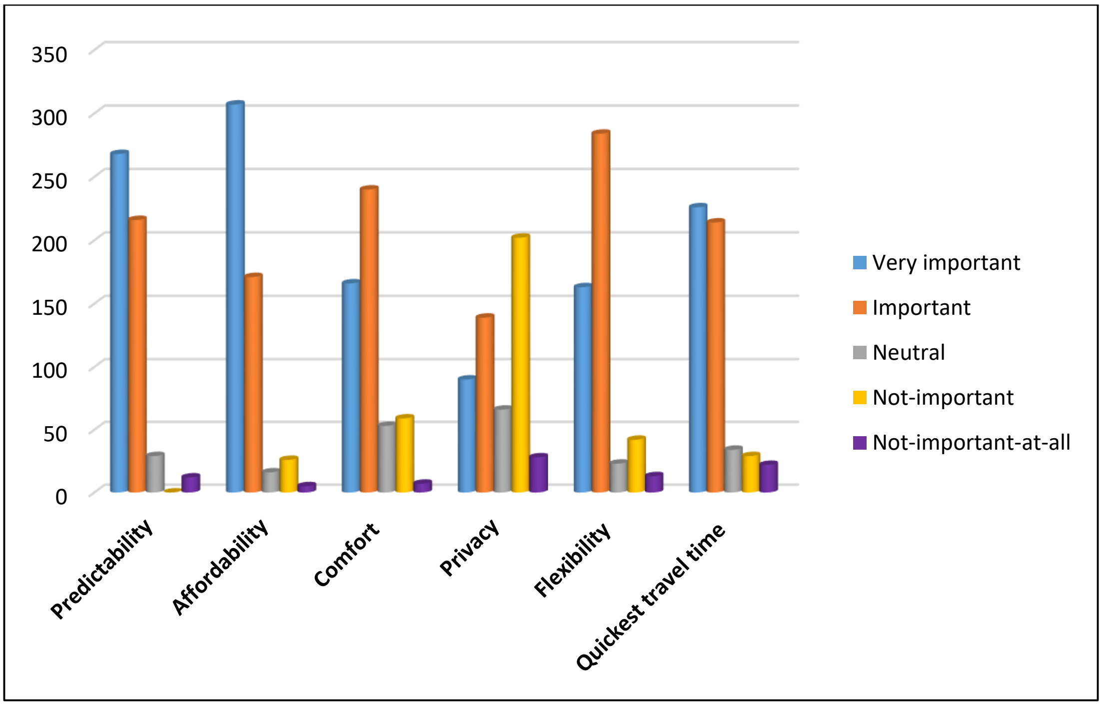


Figure 5.8: Travel mode choice considerations by public transport users
Source: Based on Field Survey, February 2015

Comfort, privacy and flexibility on the other hand, were considered very important reasons for mode choice among only 31%, 17% and 32% of public transport users. Indeed, nearly a quarter of public transport users were of the view that their travel mode did not offer comfortable commuting. Also, more than half were of the view that public transport offered less privacy. In terms of their assessment of the general condition of public transport in the metropolis, the majority (84%) were of the view that conditions were poor particularly for mini-buses/Trotro.

Furthermore, as shown in figure 5.9, among those who choose to walk to work in the metropolis, affordability and predictability were indicated as very important reasons for doing so. Whereas 25% and 47% of those who choose walking attached indicated comfort as very important and important considerations respectively, about 22% of them were indifferent. The remaining five percent were of the view that comfort was not at all an important consideration for choosing to walk to work. Moreover, about 20% of those who choose to walk expressed indifference with regards to walking being the quickest mode of transport while seven percent and three percent were of the view that this consideration was not important or not at all important reasons.

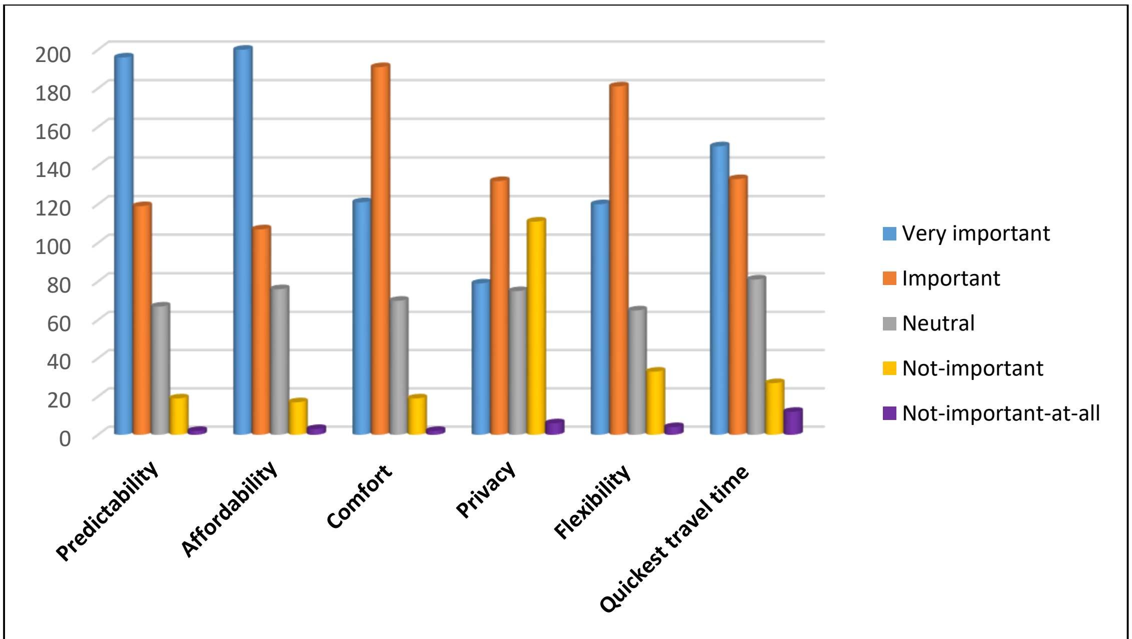


Figure 5.9: Travel mode choice considerations by individuals who walk to work
Source: Based on Field Survey, February 2015

In summary, comfort, privacy, flexibility and predictability featured prominently as reasons for mode choice among private car users. Affordability on the other hand, was not an important consideration among private car users. Among public transport users on the other hand, predictability and affordability were very important considerations while privacy and comfort featured as less important reasons for mode choice. Response of people who walk to work were nuanced as a good number of them remained indifferent to comfort, travel time, privacy as travel mode choice considerations in relation to walking.

5.6.1 Determinants of private car ownership and choice as work travel mode: a binomial logistic regression model

The private-car was the main work travel mode among 21% of workers interviewed. This section combines the choice considerations discussed above with other socio-economic and spatial variables to examine how these factors determine car ownership and choice as work travel mode using a binary logistic regression model. The analysis of private car use as work travel mode also reflect car ownership among households of the workers. In view of this, variables representing attributes of the worker's households including marital status, family size and income are included in the analysis. A summary of the variables in used in the analysis is provided in table 5.4.

Table 5.4: Variables in the binary logistic regression of private car use (N= 1158)

	Variable name	Type and coding
Dependent variable	Private car use	Categorical, coded 1-Yes 0-No
	Income-group[18]	Categorical, coded: 1- low (earnings below 25th percentile income; 2- middle (earnings between the 26th percentile and 75th percentile income; and 3- high (earnings above 75thth percentile)
	Education	Categorical coded: 1-Tertiary; 0-basic & secondary
Independent variables	Marital status	Categorical, coded: 1-couple; 0-Single
	Family size	Scale variable— Nominal values
	Distance of residence to CBD	Scale variable—of actual road distances from the home to the CBD.
	Residential zone of residence	Categorical coded: 1-Historical-core; 2-Inner-Suburb; and 3-Outer-suburb
	Average home-work distance	Scale variable—of actual road distances from the home to the work place averaged for all working members in the household
	Predictable	Categorical, code 1-important, 0-not-important
	Affordable	Categorical, code 1-important, 0-not-important
	Comfort	Categorical, code 1-important, 0-not-important
	Privacy	Categorical, code 1-important, 0-not-important
	Flexibility	Categorical, code 1-important, 0-not-important

[18] Summary statistics on the income groupings of households is presented in chapter four.

Quicker travel	Categorical, code 1-important, 0-not-important

The average distance from the place of residence to the CBD was 7km (SD = 2.85) while that of the home to the place of work was 4.5km (SD = 2.99). A hierarchical logistic regression model was specified in which the predictors were entered in a systematic manner in the order presented in table 5.4. Five models were initially formulated using this approach. The extent to which the addition of each predictor improved the model fit was assessed using the Omnibus Tests of Model Co-efficient[19]. Based on the contribution of each predictor variable to the improvement of the model, a final model was specified to explain the determinants of private car use as work travel mode. Estimates of model co-efficient are presented in table 5.5.

Table 5.5: Determinants of private-car ownership and choice as work travel mode

Predictors	*b(SE)*	95% C.I. for EXP(B) Lower	Odds Ratio	Upper
Low-income	0^b			
Middle-income	2.038 (0.61) **	2.32	7.673	25.384
High Income	3.368(0.618) ***	8.649	29.029	97.434
Education of household-head	1.491(0.187) ***	3.08	4.441	6.404
Marital status	-0.158(0.225)	0.549	0.854	1.327
Family size	0.181(0.054) **	1.079	1.199	1.332
Distance of home to CBD	-0.110(0.048) *	0.815	0.895	0.983
Historical-core residence	0^b			
Inner-suburb residence	0.7(0.303) *	1.111	2.014	3.652
Outer suburb residence	1.245 (0.369) **	1.686	3.473	7.155
Predictable	-0.456 (0.402)	0.288	0.634	1.392
Affordable	-0.542 (0.253) **	0.354	0.582	0.956
Comfort	0.265 (0.426)	0.565	1.304	3.007
Privacy	1.568 (0.255) ***	2.911	4.797	7.906
Flexibility	-0.867 (0.39) *	0.196	0.42	0.904
Quicker travel	1.385 (0.426) **	1.735	3.996	9.203
Constant	-5.677 (0.743) ***		0.003	

Note: R^2 = 0.49 (Nagelkerke), 0.33 (Cox & Snell) *$p < 0.05$, **$p < 0.01$ ***$p < 0.001$
[b]This parameter is set to zero because it is redundant

Results of the binary logistic regression showed that all the predictor variables included in the model, except marital status and the level of importance attached to comfort, had statistically significant effect on the choice of private-car as work travel mode. The likelihood of car ownership and use increases with higher household income. High income households (earning monthly income of GH¢ 2,050 and above) have odds of owning and commuting to work on a private car, 29 times higher than low income households. In the case of middle-income

[19] The initial hierarchical model specification included two models in which interaction terms were specified between (i) distance of residence to CBD and Urban-zone of residence; and (ii) marital status and family size. These did not yield improvement in model fit and were therefore excluded from the final model specification.

households (i.e. earning incomes between GH¢50 and GH¢2000), the odds are relatively smaller, although nearly eight times higher compared to low-income households. Moreover, being tertiary-educated was associated with odds of owning and commuting in a private car to work four times higher than those with low-levels of education controlling for other factors. Households with relatively larger relatively larger family size would own a private car if they could afford it. Consequently, individual workers within such households have odds of commuting to work in a private car, 1.2 times higher than those with smaller family sizes controlling for other factors.

The analysis further shows that as distance between the place of residence and the CBD decreases, the odds of a worker's household owning a car decreases. Consequently, workers living in the suburban neighbourhoods of the metropolis were more likely to commute to work in a private car than those in the historical-core neighbourhoods. Specifically, worker's having residence in the inner and outer suburban locations of the metropolis have odds of commuting to work by private car two and 3.5 times higher than those residing in the historical-core neighbourhoods.

The final part of the analysis show that the importance worker's attach to affordability, privacy, flexibility and the length of time spent in commuting to work had significant effect on private car choice. Whereas greater importance to affordability reduced the likelihood of car-ownership and hence the likelihood of individuals traveling by this mode, workers who attached greater importance to privacy were nearly five times more likely to choose the private, controlling for other factors. Similarly, workers who attached importance to flexibility in travel have odds of choosing to commute to work in a private car four times higher than those who did not.

In summary, the analyses show that socio-economic characteristics including income levels, educational attainment and family size; spatial attributes including distance between the place of residence and the CBD, urban-zone of residence (i.e. historical-core or suburban); as well as the level of importance attached to affordability, privacy, flexibility and time spent travelling determine private car ownership and choice as work travel mode. Together, these variables account for nearly 50% of the explained variance in private car ownership and use as work travel mode.

5.6.2 Determinants of choice between walking and motorized transport as work travel mode: a binomial logistic regression model

Walking constituted the primary mode of transport to work for one-third of the 1,158 individual workers interviewed. In view of this, the analysis examines the factors that determine whether an individual worker would choose walking over motorized forms of transport[20]. Table 5.6 shows the variables used in the analysis.

Table 5.6: Variables in the logistic regression of choice between walking and motorized transport (N= 1150)

	Variable name	Type and coding
Dependent variable	Walking vs. Motorized transport	Categorical, coded: 1-Walking 0-Motorized (private car, public transport)
Independent variables	Income-group	Categorical, coded: 1- low (earnings below 25th percentile income; 2- middle (earnings between the 26th percentile and 75th percentile income; and 3- high (earnings above 75thth percentile)
	Education	Categorical coded: 1-Tertiary; 0-basic & secondary
	Marital status	Categorical, coded: 1-couple; 0-Single
	Non-car-ownership	Categorical, coded 1-Yes 0-No
	Home-work distance	Scale variable— actual road distances from the home to the work place averaged for all working members in the household
	Job location	Categorical coded: 1-home-based; 0-non-home-based
	Affordable	Categorical, code 1-important, 0-not-important
	Flexibility	Categorical, code 1-important, 0-not-important
	Quicker travel	Categorical, code 1-important, 0-not-important

A hierarchical approach was adopted in which separate models were specified for each of the predictor variables in the model in the order presented in table 5.6. The extent to which the addition of each predictor improved the model fit was assessed using the Omnibus Tests of Model Co-efficient. One of the initial models included urban-zone of residence as predictor variable and an interaction term between that variable and home-work distance. These did not contribute to the model fit and where therefore removed from the final model specification. The model estimates of the likelihood of choice of walking over motorized transport to work is presented in table 5.7.

[20] The analysis therefore, is based on data from 1150 out of the 1158 individuals interviewed; motorbike and cycling as work travel modes have been excluded from the analysis due to low data counts.

The results show that controlling for other factors, the likelihood of individual workers choosing to walk over motorized transport to work decreases as income increases[21]. Workers from low-income households (i.e. households earning less than GH¢ 750 a month) have odds of walking to work, five times higher than workers from high income households (i.e. households earning monthly income of GH¢ 2,050 and above). Also, workers from lower-middle income households have odds of walking to work almost four times higher than those from high income households. Besides income levels, non-tertiary educated workers with odds ratio of 2.08 were twice more likely to choose walking over motorized transport as work travel mode.

Table 5.7: Determinants of choice between walking and motorized transport as work travel modes

		95% C.I. for EXP(B)		
	b(SE)	**Lower**	**Odds Ratio**	**Upper**
High-Income	0[b]			
Low-income	1.691 (0.342) ***	2.776	5.425	10.6
Middle-income	1.299(0.294) ***	2.062	3.666	6.518
Education	0.734(0.262) **	1.248	2.083	3.478
Non-car-ownership	1.529(0.334) ***	2.397	4.613	8.878
Job location (home-based)	2.838(0.391) ***	7.936	17.089	36.801
Home-work distance (km)	-0.37(0.055) ***	0.621	0.691	0.769
Home-work distance * Job location (home-based)	0.181(0.085) *	1.014	1.198	1.415
Affordable	0.969(0.279) **	1.524	2.634	4.553
Quick	-0.385(0.3)	0.378	0.68	1.226
Flexibility	0.165(0.284) *	0.676	1.18	2.06
Constant	-3.227(0.471) ***		0.04	10.6

Note: R^2 = 0.65 (Nagelkerke), 0.46 (Cox & Snell) *p < 0.05, **p < 0.01 ***p < 0.001
[b]This parameter is set to zero because it is redundant

Moreover, the lack of a private car in the household increases the odds of individual working members of that household walking to work: controlling for other factors, workers belonging to non-car-owning households have odds of walking to work nearly five times higher than those with households owning a private car. Workers are more likely to drive or use public transport to work rather than walk as home-work distance separation increases and vice versa. Indeed, the results show that workers whose work-place was located inside the home or within the immediate vicinity of the home (i.e. home-based workers) have odds of walking to work 17 times higher than non-home based workers. The interaction terms specified between whether an individual's job location was home-based or non-home-based, and the road distance separation between the home and work-place also lends credence to the above findings.

[21] Since income as a predictor variable has three categories, high-income has been set as the reference category against which the likelihood of walking to work for individuals belonging to low and middle income categories is compared.

Generally, the analysis found that home-work distance separation exceeding 0.3km would be accessed using the vehicle while walking was associated with relatively shorter distances.

Finally, workers who attached greater importance to affordability in choosing their work travel mode, were nearly three times more likely to walk to work than choose a motorized form of transport, controlling for other factors. While the need for the quickest mode of travel decreased the odds of walking, workers who prioritized flexibility in their work travel mode choice were a little more likely to walk to work than opt for motorized transport (odds ratio = 1.18).

5.6.3 A composite logistic regression model of work transport mode choice among non-private car users

The analyses of transport mode use to work have so far in the preceding sections, examined the determinants of private car ownership as well as the factors influencing the choice between walking and all forms of motorized transport. This section specifies a multinomial logistic regression model that combines choice between three non-private transport modes—Mini-buses/Trotro and Taxis as public transport mode alternatives and walking. The data used for this analysis excludes private-car ownership and use as the preceding analysis has established car owners consistently commute to work in their private cars. Also, bicycle and motorcycle use are excluded from the analysis due to low representation of these modes in the data. The analysis is therefore based on data from 928 individuals out of the total 1,158 interviewed.

Results of the analysis is presented in Table 5.8. Walking, the third category of transport mode, was set as the reference category in the model. This means that the model estimates the likelihood of a worker choosing to commute to work in a Mini-bus/Trotro as compared to walking in the first instance, choosing to commute by a Taxi as compared to walking in the second instance and finally, choosing between Mini-bus/Trotro and Taxi.

The comparison of choice between mini-bus as a public transport mode and walking to work shows that the odds of using the former is 1.3 times higher than using the latter as distance increases beyond a quarter of a kilometre. In the case of choice between taxi and walking, the odds were 1.5 times in favour of the former as home-work distance exceeds a quarter of a kilometre. In the case of choice between the Mini-bus/Trotro and Taxi, the analysis show that increasing home-work distance decreased the likelihood of selecting the later over the former, controlling for other factors.

Table 5.8: Determinants of choice between walking, and two public transport mode alternatives (N= 928)

		95% Confidence Interval for Exp(B)		
	b (SE)	**Lower Bound**	**Odds Ratio**	**Upper Bound**
Public transport (Mini-bus/Trotro) vs Walking				
Intercept	-4.784 (0.503) ***			
Home-work distance (km)	0.292 (0.045) ***	1.227	1.339	1.461
High Income	1.565 (0.362) ***	2.351	4.782	9.726
Middle-income	0.388 (0.253)	0.897	1.473	2.42
Low-income	0^b			
Home-based job location	3.535 (0.344) ***	17.495	34.308	67.278
Non-Home-based job location	0^b			
Affordability *(not-important)*	1.994 (0.362) ***	3.616	7.345	14.919
Affordability *(important)*	0^b			
Comfort *(not-important)*	-0.542 (0.347)	0.295	0.581	1.147
Comfort*(important)*	0^b			
Privacy *(not-important)*	-1.057 (0.236) ***	0.219	0.347	0.552
Privacy (*important)*	0^b			
Quick *(not-important)*	-0.172 (0.313)	0.456	0.842	1.555
Quick (*important)*	0^b			
Public transport (Taxi) vs Walking				
Intercept	-6.6 (0.85) ***			
Home-work distance (km)	0.297 (0.056) ***	1.206	1.345	1.501
High Income	1.788 (0.541) **	2.071	5.977	17.249
Middle-income	1.005 (0.454) *	1.121	2.732	6.657
Low-income	0^b			
Home-based job location	2.346 (0.503) ***	3.899	10.445	27.984
Non-Home-based job location	0^b			
Affordability *(not-important)*	-0.297 (0.412)	0.331	0.743	1.667
Affordability *(important)*	0^b			
Comfort *(not-important)*	0.731 (0.666)	0.563	2.078	7.669
Comfort*(important)*	0^b			
Privacy *(not-important)*	0.871 (0.397) *	1.097	2.388	5.199
Privacy (*important)*	0^b			
Quick *(not-important)*	0.319 (0.565)	0.455	1.376	4.161
Quick (*important)*	0^b			.
Mini-bus/Trotro vs Taxi				
Intercept	0.415 (0.703)			
Home-work distance (km)	-0.005(0.044)	0.912	0.995	1.086
High Income	-0.223(0.509)	0.295	0.8	2.171
Middle-income	-0.617(0.452)	0.223	0.539	1.307
Low-income	0^b			
Non-home-based job location	1.189(0.567) *	1.081	3.285	9.976
Home-based job location	0^b			
Affordability *(important)*	-2.291(0.433) ***	0.043	0.101	0.236
Affordability *(not-important)*	0^b			
Comfort *(important)*	1.274(0.64) *	1.02	3.573	12.517
Comfort*(not-important)*	0^b			
Privacy *(important)*	1.928(0.371) ***	3.325	6.874	14.211
Privacy *(not-important)*	0^b	.	.	.
Quick *(important)*	0.491 (0.544)	0.562	1.634	4.745
Quick *(not-important)*	0^b			

Note: R^2 = 0. 60 (Nagelkerke), 0.51 (Cox and Snell) *p < 0.05, **p < 0.01 ***p < 0.001
[b]This parameter is set to zero because it is redundant

Furthermore, controlling for other factors, high-income earners who commuted to work in public transport, have odds of using the Mini-bus/Trotro, five times higher than walking to work: Between walking and commuting to work in a taxi, the odds is higher at almost six times in favour of choosing a taxi among workers from high income households. High-income levels also decreased the odds of choosing Mini-bus/Trotro over Taxi as work travel mode (odds ratio = 0.8). In the case of workers from middle-income households, the odds of choosing Mini-bus/Trotro walking was one and a half times higher than walking to work. Middle-income earners were more likely to choose the Taxi over walking (odds ratio = 2.73) and less likely to choose the Mini-bus/Trotro over Taxi (odds ratio = 0.54).

Thus, overall, as income increases, households who do not have private cars tend to travel to work in one of two public transport modes—mini-bus/Trotro or taxi. Even so, between the two modes, high income earners have stronger preference for the taxi than the Mini-bus/Trotro. Indeed, likelihood estimates for a separate binary choice analysis specified between the Mini-bus and taxi showed that high-income earners have odds of commuting in a Taxi 3.5 times higher than commuting in a Mini-bus/Trotro (B = 1.256, Wald = 0.024; 95% C.I. for EXP(B): lower bound = 1.180, upper- bound = 10.45).

Moreover, home-based job location was associated with odds of choosing to walk to work, 34 times higher than doing so in a mini-bus/Trotro. Between the taxi and walking, home-based workers also had greater likelihood of choosing the latter over the former. Non-home-based workers were also three times more likely to choose the Mini-bus/Trotro over Taxi as work travel mode controlling for other factors.

Finally, the importance attached to affordability in work travel mode choice increases the likelihood of choosing the Mini-bus/Trotro over walking. This finding might appear contradictory. However, given that most non-home-based workers in the lower-income brackets commute by Mini-bus/Trotro and prioritise affordability, the effect of their preference would impact more on the analysis compared to home-based workers most of whom choose to walk to work because their jobs are in the home or within walking distance of the home where incurring no travel cost is a given. On the contrary, between Mini-bus/Trotro and Taxi, workers who prioritize affordability were less likely to choose the latter given the relatively higher fares associated with travelling in a Taxi. In addition, commuters who prioritize privacy in the choice of work travel mode were nearly seven times more likely to choose the Taxi compared to the Mini-bus/Trotro.

5.7 Analysis of work commuting times and costs

This section analyses data on individual's travel times between their home-origin and work-destinations as well as the out-of-pocket cost associated with work commute for the different motorized modes (i.e. private and public transport). The key factors determining work travel times and transport costs are also examined.

5.7.1 Commuting times to and from work

For this analysis, data was obtained from workers on work journey start times and return times for a five-day period. Based on this data, the morning home-to-work trip start times and evening work-to-home trip start times (see figure 5.10) as well as the actual commuting times for each leg of the trips were derived.

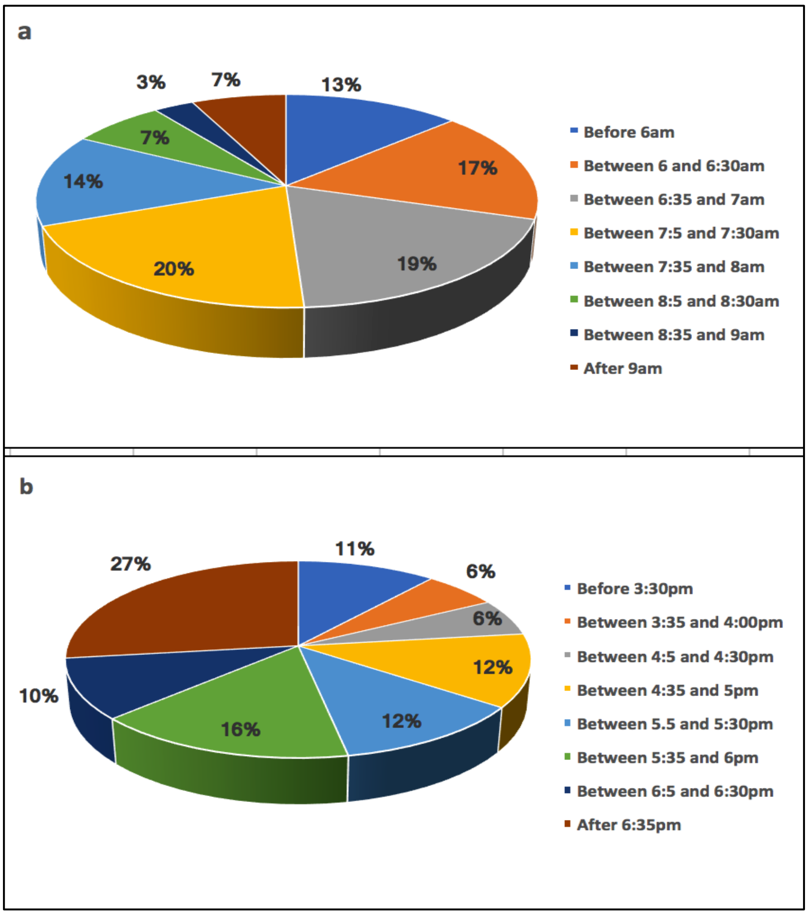


Figure 5.10: Start times for (a) Morning home-to-work trips and (b) Evening work-to-home trip
Source: Based on Field Survey, February 2015

The analysis show that a significant proportion of morning home-to-work travel (80%) occur between the early hours of 6:00am and 9:00am (figure 5.10a)[22]. Whereas, nearly half of all work trips start times within this period took place in the early morning off-peak period between 6:00 am and 7:00 am, 44% of them were taken during the early morning peak hours of commuting, between 7 and 9 am. The remaining seven percent of morning work trips occurred during the off-peak periods after 9am. In the case of return journeys (figure 5.9b), whereas about 23% of trips were taken during the off-peak hours before 4:30pm, about 40% of evening work-to-home trips occurred during the peak hours between 4:30pm and 6:30pm. The remaining 37% of journeys occurred during the evening off-peak hours after 6:30pm.

Furthermore, the analysis revealed a positive association between morning work commute start times and whether the goal was to escape traffic congestion or not (Pearson $X^2 = 35.331$, df = 6, $p < 0.00$, Cramer's V = 0.20) and that of return work trip times in the evening (Pearson $X^2 = 26.241$, df = 8, $p < 0.00$, Cramer's V = 0.14)[23]. Overall, about 49% of the respondents, who commuted to work using motorized transportation (i.e. public and private transport) indicated that they chose the times they went to work in the mornings to avoid traffic congestion. Among morning off-peak commuters 54% and 22% of journeys starting between 7 and 9am and after 9am respectively, were taken by commuters with the with the goal to escape traffic congestion. In the case of return journeys in the evening, 28% of all commuters indicated that they chose their traveling times with the goal to escape traffic congestion. Within this group of commuters, 29% and 30% of trips that occurred during evening off-peak periods before 4:30pm and after 6:30pm respectively, were taken with the goal of avoiding traffic congestion.

Thus, most of the morning peak-hour commuters (54%) and evening peak-hour commuters (74%), did not choose their travelling times to avoid congestion. Instead, their travel times in the mornings were scheduled with the goal to arrive at the work-place at stipulated reporting times while evening travel times started after stipulated work closing times. Return work journeys were therefore consistent with formal rules regarding work reporting and closing hours, which act as a form of authority constraint on travel times. The observed peak-hour

[22] Work commute start times were found to be consistent among the respondents. Consequently, Individuals who commuted at certain times of the day three or more times during the five-day period, were assigned that time as their regular work trip start times in the mornings and in the afternoon.
[23] Work travel times were divided into three categories namely; off-peak, peak, and off-peak hours for both morning and evening trips; goal of avoiding traffic was dichotomized as either yes or no.

travel, characterized by congestion is therefore a consequence of constraints imposed by these formal rules rather than the result of the personal choices of individual commuters regarding when they decide to travel.

Summary descriptive statistics of data on actual times spent traveling to and from work are presented in table 5.9. The analysis show morning and evening work travel times for all trips, trips associated with home-based work only and trips associated with non-home-based work only.

Table 5.9: Travel times for morning work trips

Travel time (Minutes)	**All single work trips (%)**		**Home-based work trips only (%)**		**Non-home-based work trips only (%)**	
	Home-work (mornings)	Work-home (Evenings)	Home-work (mornings)	Work-home (Evenings)	Home-work (mornings)	Work-home (Evenings)
below 10 minutes	19	11	58	40	6	2
10 to 19	21	5	27	14	18	2
20 to 29	17	2	9	4	19	1.5
30 to 39	18	67	2	36	24	77
40 to 49	12	1	1	0	16	1
50 to 59	4	1	2	0	5	1.5
60+	10	12	0	6	13	5
Average	25.0	32.85	5.0	10	30.0	32.86
SD	19.79	18.61	10.89	19.37	19.01	16.131
Maximum	120	105	60	105	120	105

Source: Based on Field Survey, February 2015

On the average, commuters spent 25 minutes and 33 minutes travelling from home-to-work in the mornings and returning to home from work in the evenings respectively. For home-based work (work located within the immediate vicinity of the dwelling) where trips were completed by walking, the average walking time to work was 5 minutes and 10 minutes in the mornings and evenings respectively. The exact reasons for the difference in walking times in the morning and evening could not be inferred directly from the survey. Notwithstanding, it is possible that individuals might take relatively longer routes where perhaps there might be others using it or there might be good lighting in evening for safety reasons. In the case of trips associated with non-home based work, the average travel times was 30 minutes and 33 minutes in the mornings and evenings respectively.

Aggregating the travel times for return trips associated with home-based work and non-home-based work, individuals spent on the average an hour (SD = 32.32, maximum = 180 minutes) travelling to and from work daily. In the case of trips associated with home-based work only, individuals spent about 27 minutes (SD = 25.053, maximum = 120 minutes) daily travelling to and from work while non-home-based workers spent a little above an hour (62 minutes) doing same (SD = 27.30, maximum = 180 minutes).

Spearman's Rho correlation analysis shown a positive association between job location (i.e. whether home-based or non-home-based) and travel time ($r = 0.581$, $p < 0.001$); mode of transport to work and travel time ($r = 0.634$, $p < 0.001$)[24] and TAZ origin and destination of work trips and travel time [25] ($r = 0.547$, $p < 0.001$). A positive association was also found between home-work distance and travel time ($r = 0.597$, $p < 0.001$). Based on the above relationships, a linear regression model was fitted to quantify the extent to which these factors determined total work travel time (see table 5.10)

Table 5.10: Determinants of work travel times

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Constant	-26.220	3.648		-7.188	.000
Travel-mode (walk vs motorized)	20.605	2.189	.302	9.412	.000
Travel mode*Distance	.326	.110	.092	2.959	.003
Work location (home-based vs non-home-based)	17.371	2.278	.232	7.627	.000
TAZ Origin and Destination	11.631	1.930	.179	6.026	.000

Note: $R^2 = 0.45$

The results of the analysis show that using motorized transport was associated with seven-minute increase in travel time. This is reasonable since as indicated earlier, most non-home-based work travel, which are completed using vehicles also tend to be longer compared to home-based work trips which involve relatively shorter or no travel distance and are mainly completed as walking journeys. Indeed, non-home-based work location increases time of commute by about seven minutes. Controlling for other factors, the effect of the interaction term specified between travel mode and travel distance, though statistically significant is almost negligible. Finally, travel time increases by almost 11 minutes if one's trip origin TAZ is different from their trip destination TAZ. The model could account for 45% of the variance in commuting times.

[24] Transport mode was dichotomized into two categories; (i) non-motorized—walking) and; (ii) motorized—private car and public transport

[25] TAZ origin and destination of trips was dichotomized into two categories; (ii) whether TAZ origin and destination of trips is the same; and (ii) whether TAZ origin and destination of trips is different.

5.7.2 Analysis of work travel costs

This section analyses data obtained on individual workers' expenditure on transport, representing out-of-pocket costs incurred as fares for public transport users and fuel cost for private car users. Using this information, monthly transport cost as a proportion of household income was assessed for private car users and public transport users as summarised in table 5.11.

Table 5.11: Total monthly out-of-pocket transport costs for different income groups and mode use

Travel Mode	Income-group	Percentage	Monthly Income Range (GH¢)	Monthly Transport Cost(GH¢)		Transport cost as percentage of income
				Median	SD	
Private car users [26]	Low	1	150 -700	100	141.0	66 -14.0
	Lower-middle	11	750 - 1200	175	117.6	23 -14.5
	Upper-middle	27	1250 - 2000	120	138.0	9.6 - 6.0
	High	47	2050 - 4000	275	265.8	13.4 - 6.9
	Rich	14	4100 - 14000	300	338.0	7.3 - 2.1
Public transport users	Low	19	150 -700	50	59.1	33.3 - 7.2
	Lower-middle	33	750 - 1200	63	40.2	8.4 - 5.2
	Upper-middle	27	1250 - 2000	100	72.6	8.0 - 5.0
	High	19	2050 - 4000	139	94.0	6.8 - 3.5
	Rich	2	4100 - 14000	250	146.5	6.1 - 1.8

Source: Based on Field Survey, February 2015

As expected, private car users spent more (GH¢200) on the average as fuel costs compared to the average monthly transport fare cost of GH¢80 for public transport users. Overall, the amount of income spent on transport increases with higher incomes. However, transport cost as a proportion of household income decreases as income increases. For example, whereas workers from low income households who use private cars spent on the average GH¢100 every month on transport, representing between 14% and 66% of their incomes, workers in households classified as rich spent on the average GH¢300 every month on transport representing between two and seven percent of their monthly earnings.

In addition to the total monthly spending on transport, the factors influencing transport costs among workers are examined. An initial Spearman Rho correlation analysis revealed a positive association between household income and transport cost ($r = 0.548$ $p < 0.001$) and negative relationship between mode of transport (i.e. private vs public transport) and travel costs ($r = -0.538$ $p < 0.001$). Furthermore, Pearson correlation analysis showed a rather small but significant negative association between transport costs and number of working days (r = -

[26] The analysis is based on fuel costs only and does not include other cost such as car maintenance and insurance premiums.

0.113 p = 0.003); positive association between distance travelled to and from work and transport cost (r =0.123 P < 0.001) and a positive relationship between number of adult workers in a household and total transport spending (r = 0.245, p <0.001).

Based on the above relationships, a linear regression model was specified to quantify the relationship between these factors and commuting costs. Results of the analysis is presented in table 5.12. It shows that total out-of-pocket spending on commuting increases by GH¢30.981 as the number of adult working members in a household increases by one, controlling for other factors. Quite unexpectedly, household spending on commuting tend to decrease by GH¢13.85 as the number of working days in a week increases by one. The mode choice for work and type of work could explain this. Public transport users, who constitute the largest share of motorized commuters, on the average worked six times in a week while private car users on the average worked a day less in a week. Given that transport cost is generally higher for private car users and lower for public transport users, the decrease in transport cost as the number of working days increase also seem reasonable. Indeed, as the results of the regression analysis show, private car use, increases total monthly household spending on transport by GH¢36.99.

Table 5. 12: Determinants of work travel costs[27].

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	90.168	46.737		1.929	.054
Number of workers in household	30.981	5.207	.258	5.950	.000
Work days (trip frequency)	-13.851	3.836	-.142	-3.611	.000
Home-work travel distance	.667	.381	.071	1.753	.080
Work location (home vs non-home based)	26.473	16.762	.064	1.579	.115
Private Car Ownership	36.986	12.995	.117	2.846	.005
Public Transport mode (taxi vs mini-bus)	-47.241	8.515	-.220	-5.548	.000
Income group	27.182	5.575	.211	4.876	.000

Note: $R^2 = 0.31$

Moreover, among public transport users, households whose working members choose to travel to work in a Mini-bus/Trotro spent GH¢47.241 less than those who commute to work using Taxi. As household income increases from low (i.e. GH¢150 - GH¢700) to high (i.e. GH¢4100 - GH¢14000), total monthly work travel spending increases by GH¢27.18 controlling for other

[27]See appendix 1 for scatter plots of the relationship between the variables in the model. The scatter plots show that the data is rather noisy, as such, the relationships among the variables are only weakly expressed in the regression analysis.

factors. This is because car ownership, which is associated with relatively higher spending, is common among households of relatively higher incomes. About 31% of the variance in households' transport spending is accounted for by the variables included in the analysis.

5.8 Chapter summary

This chapter set out to provide empirical analysis of mobility characteristics in the Kumasi metropolis, the case study for this research based on travel data elicited from some 1,158 randomly sampled individuals. The analyses focused on four key aspects of mobility namely; home-work travel production and attraction patterns, work travel mode choice, travel times and commuting costs.

Anchoring the existing TAZ system of the metropolis to the spatial distribution of key metropolitan land-use function provided an understanding of the urban structural conditions within which commuting for work purposes occurred. The analysis of home-work trip production and attraction patterns in the form of origin and destination flows at the level of aggregate TAZs reflected the location of major urban functions such as industry, retail and service, education and residential. For example, while one-third of work trips had home-origins in the two most central TAZs, half of all work trips terminated in these zones, reinforcing the dominant commercial and service functions of these zones. A large proportion of all work trips were generated and distributed among the six internal zones of metropolis. Moreover, nearly half of work trips started and ended within the same TAZ, signifying a relatively shorter distance separation between individual's work and home locations. This was reinforced by the finding that about one-third of jobs were home-based requiring less or no travel at all while among non-home-based workers, the average home-work distance separation was 4.5km.

Furthermore, the determinants of work travel mode choice were examined using a series of logistic regression models. For private car ownership and use, the analysis found that socio-economic variables including household income, educational levels and family size; spatial variables such as home-CBD distance separation and residence in suburban areas; and importance attached to privacy, comfort and travel speed/time were the key determinants. Home-based job location, residence in the historical-core neighbourhoods, home-work job distance, priority for affordability, household income and education levels of commuters were the key determinants of choice between non-motorized (i.e. walking) and motorized transport (i.e. private and public), and between public transport modes (i.e. Minibus/Trotro and Taxi).

The final dimension of the analysis of mobility patterns examined the determinants of work travel times and costs. Among the key findings were that commuting times were influenced by work travel mode choice, distance separation between the home and the work-place, work location (i.e. home-based and non-home-based) and the TAZ origin and destinations of work trips. In terms of commuting costs, the analyses revealed that household spending on transport was determined by whether multiple workers were present, the number of working days in a week, car ownership, whether public transport users commuted to work using taxi or mini-bus/Trotro and the income levels of households. Overall, transport cost as a proportion of household income, decreases with increasing income implying that in percentage terms, low income households spent more of their income on transport compared to households of higher means.

The empirical analysis of mobility patterns presented in this chapter, together with the analysis of residential and job location choice presented in chapter four, have offered critical snapshot understanding of the relationship between long-term urban location choice decisions and short-term choices related to daily patterns of mobility. The empirical insights accrued will provide the foundation to operationalize the second objective of this research, which seeks to develop a dynamic simulation model of the co-evolution of urban location choice and mobility patterns. The model development process begins next in chapter six.

CHAPTER SIX: SPECIFICATION OF A METHODOLOGICAL FRAMEWORK TO SIMULATE URBAN LOCATION CHOICE AND MOBILITY PATTERNS

6.1 Introduction

In Chapter two, two main objectives of this thesis were outlined. The first objective was to understand empirically, urban location choice and mobility patterns in the Kumasi metropolis, the case study area for this research. Following from this objective, results of two empirical studies were presented in Chapters Four and Five. The second objective outlined for this thesis was to develop a disaggregate model to simulate the urban location choice process and the associated patterns of mobility in the case study metropolis. It was explained that these research objectives are mutually linked in that the first was deployed to gather critical empirical insights that would provide the basis for the implementation of the model indicated by the second.

The goal of this chapter, is to put forward a methodological framework upon which the disaggregate model can be implemented. The methodology presented here therefore formulates a conceptual model based on a novel modelling paradigm that integrates theories, principles and concepts regarding how location choice co-evolve with patterns of spatial interaction observed in urban areas.

The specification of the conceptual model follows a systematic approach that begins with a brief review of existing modelling paradigms. Here, two main urban modelling traditions—aggregate and disaggregate modelling are reviewed highlighting their strengths and limitations. Following this, the most suitable modelling approach is selected based on considerations including the model purpose, innovation and the overall direction of research in the field. Using the selected modelling approach, a detailed description of the conceptual model is presented in the final step.

It is worth mentioning that, given its methodological focus, this chapter does not provide detailed discussion of the model implementation process. Instead, the conceptual model specification sets out the modelling paradigm to be adopted and the purpose and overall structure of the model, identifying the model design principles, entities, variables and parameters. A detailed discussion of how the conceptual model has been operationalized will follow in Chapter seven.

6.2 Chapter organization

The remainder of this chapter is structured into four sections. In the first section, a brief review of the two main traditions in urban modelling—aggregate and disaggregate approaches—is presented, highlighting their relative strengths and weaknesses. Following from this, an argument is put forward to justify the selection of a hybrid agent-based and geospatial modelling approach as the most suitable and novel paradigm for the development of the model. The third section utilizes the ABM approach to formulate a conceptual model in which the purpose, structure and design concepts are outlined. The concluding section presents a summary of the chapter and points to the linkages between this chapter and the subsequent chapters.

6.3 Brief evaluation of urban modelling approaches

Urban modelling approaches have over the years, co-evolved with theory development as well as advances in computing and geographical information systems (Wise et al., 2016). Consequently, since the early 1970s, there has been a gradual shift from aggregate, deterministic urban models grounded in classical econometric and entropy-based spatial interaction modelling traditions towards disaggregate stochastic modelling approaches broadly classified in the literature as micro-simulation models. Several theoretical propositions including Random-utility theory (McFadden, 1973), Time-geography theory (Hagerstrand, 1970; Chapin, 1974) as well as Systems and Complexity theory (Batty, 2007; Allen, 2012; Moroni, 2015; Forrester, 1993) have driven the transitions towards disaggregate modelling.

The literature on disaggregate modelling is extensive. In Chapter two, the theoretical foundations of this modelling paradigm, the accompanying techniques as well as models that have emerged from it were discussed. Here, a summary of the relative strengths and limitations of aggregate and disaggregate urban modelling is presented.

As summarized in table 6.1, the disaggregate modelling paradigm has several advantages over aggregate approaches. This includes the ability to represent a wide range of urban actors or decision-makers (i.e. heterogeneous actors) and their complex behaviours which derive from their personal-level attributes as well as the attributes of their surrounding environments (Pinjari and Bhat 2011; Rasouli and Timmermans 2014). Moreover, disaggregate urban modelling approaches helps to overcome and/or relax weak assumptions including spatial homogeneity, monocentricity, spatial equilibrium as well as the presence of unboundedly rational decision-makers with unlimited access to information (Batty, 2017; Acheampong and Silva, 2015).

Table 6.1: Comparison between aggregate and disaggregate urban modelling approaches

Aggregate modelling approach	Disaggregate modelling approach
▪ Homogenous urban actors/decision-makers	▪ Representation of heterogeneous urban actors/decision makers
▪ Representation of urban landscape as homogenous	▪ Representation of urban landscape as heterogeneous
▪ Simplifying and restrictive assumptions including rationality, unlimited information flow and access	▪ Ability to overcome or relax simplifying assumptions and improve model realism by incorporating bounded rationality, heuristics and choice situations under conditions of uncertainty
▪ Assumption of urban systems in equilibrium	▪ Urban systems are assumed to be in perpetual disequilibrium
▪ Often top-down deterministic specification of urban systems and behavioural rules	▪ Often bottom-up specification of complex and adaptive rules leading to self-organization and emergent behaviour.
▪ Simple, tractable and requires less computational power	▪ Complex, high computational demand, stochastic variation and output uncertainty

6.3.1 Agent-based modelling as a disaggregate urban modelling paradigm

ABM as a disaggregate urban modelling paradigm, has it origins in the micro-simulation approaches to urban modelling. Fundamentally, the concept of micro-simulation is one in which the aggregate behaviour of a system is explicitly simulated over time as the sum of the actions and interactions of disaggregate behavioural units within the system (Iacono, et al., 2008; Miller and Savini, 1998). Micro-simulation models in general, derive their strength from their dynamic nature, which makes it possible to trace model components (e.g., individuals, households, jobs, and dwellings) over time and to observe the modelled processes of change at a level of detail that is not possible in other types of models (Pagliara and Wilson, 2010; Huang et al., 2014).

ABM is a bottom-up computational method that allows for the creation, analysis and experimentation with models composed of autonomous agents that interact with each other and their environment locally (Gilbert 2008, Railsback and Grimm 2011, Railsback et al., 2006, Silva, 2011). Rooted in general systems theory and complexity theory, ABM as a microscopic modelling paradigm allows for a natural description of a complex system in a flexible and robust manner to capture emergent phenomenon (Batty 2001, Bonabeau 2002, Castle and Crooks 2006, Wu and Silva 2010). The approach allows one to represent heterogeneous agents (e.g., household members or individuals within the simulated population) who can learn, modify, and improve their interactions with their environment (Batty 2007, Pinjari and Bhat 2010, Jin and White 2012,). As an emerging methodology that continues to find new applications in different disciplines, the field of ABM has become established as one of the innovative approaches to represent and simulate multi-scale urban dynamics (Batty, 2005; Arsanjani et al., 2013; Martinez and Morales, 2012; Zhang et al., 2010; Parker et al., 2003; Bithell and Brasington, 2009).

The conceptual model presented in this chapter therefore adopts the ABM approach because of the inherent capabilities and advantages of the technique, and in line with the overall direction of current research in the field of urban land use and travel behaviour modelling. Moreover, the choice of ABM derives naturally from the objective of this research to develop a disaggregate model to simulate residential-job location choice behaviour and the associated mobility patterns of heterogeneous urban households. Such a model, ought to be able to represent the complex micro-level behaviour of individuals' choice decisions and property market dynamics over time in a spatially explicit framework. ABM simulation paradigm makes this possible by allowing for a bottom-up representation of preferences of heterogeneous agents (i.e. households) for spatially diverse goods (i.e. dwelling units, land parcels and employment locations), the idiosyncratic differences in decision-making processes as well as important feedback relationships (Magliocca et al., 2011; Murray-Rust et al, 2013; Huang et al., 2014). In addition, ABM has proved useful in representing bilateral transactions and competition among different actors in housing and property markets as the basis for demand and supply dynamics and price determination (see e.g. Ettema, 2011; Waddell 2001; Parker and Filatova, 2008; Filatova et al., 2011).

Furthermore, the approach provides a platform to hybridize the strengths of standard urban econometrics with the principles of complexity theory and bounded rationality/decision making under uncertainty. Consequently, the unrealistic assumptions of aggregate agent behaviour, spatial homogeneity and systems equilibrium under conditions of rationality and perfect information can either be relaxed or completely overcome in ABM models of complex systems (Batty, 2008; Manson, 2006; Rasouli and Timmermans 2014a; Filatova et al., 2011).

Finally, ABM allows for the integration of a wide range of data sources including expert knowledge, empirical research, insights from existing models, survey data and geospatial data to model complex urban processes. Models could also be adapted to different contexts through the modification of model parameters and calibration based on the prevailing realities of the context (Murray-Rust et al., 2013; Janssen and Ostrom, 2006; Robinson et al., 2007).

6.4 Specification of the conceptual model

In this section, a conceptual model to simulate how urban location choice co-emerges with mobility patterns is specified using ABM paradigm. Disaggregate models adopting the ABM approach are complex due to the many interacting components which individually have several entities, parameters and state variables. For comprehensive, logical and easy to understand description of such models, the ODD protocol—Overview, Design Concept and Details, has been proposed as a standardized model description procedure (Railsback and Grimm, 2012).

The first element of the protocol—Overview states the purpose of the model and the overall structure, specifying the entities, state variables and spatial scales involved. Under the second element of the protocol—Design, the basic design concepts underlying the model including sensing, learning, adaptive behaviour, stochasticity, interaction, feedback and emergence are discussed. The final element of the protocol—Details, provides an overview of the decision-making framework used by agents in the model to accomplish various decision-making tasks implemented in the model. Following the protocol strictly requires that the last element also include a detailed description of how the model is initialized, how data are processed and inputted as well as all the major processes in the model. However, because the aim is to specify a conceptual model at this stage, such details are postponed to Chapter seven where the actual implementation of the model is discussed.

6.4.1 Overview of the conceptual model

- Purpose of the Model

The purpose of the model is to simulate how urban spatial structure and mobility patterns co-evolve as a function of the interaction between the residential and job location choice behaviour of heterogeneous households and individuals in the urban property and job markets and existing urban structural conditions.

- Model Structure: sub-components, entities and variables

The overall structure of the conceptual model is illustrated in figure 6.1. The framework comprises six main interacting model sub-components. The entities, key variables and data required are outlined under each of the sub-components. In the sections that follow, a description of the individual sub-components of the framework is presented.

I. Spatio-environmental sub-component

The spatio-environmental sub-component represents the geographical context of the model. As shown in figure 6.1, this sub-component is divided into two main elements namely; the urban spatial or physical units of analysis and the urban functional attributes. The first elements comprise macro-level geo-spatial input data such the metropolitan boundary and existing administrative sub-divisions.

The conceptual model recognizes the need for further spatial differentiation beyond the administrative boundaries. Consequently, three broad urban-zones (U_Z), of unique socio-spatial characteristics namely; Historical-core (U_{Z1}), Inner-suburb (U_{Z2}) and Outer-suburb (U_{Z3}) are identified as meso-scale spatial units for anchoring the model input data during implementation. The second element of the spatio-environmental sub-component is the urban functional attributes in the metropolitan context. Thus, at the implementation stage of the conceptual model, the spatio-environmental sub-component takes as inputs, real-world geo-spatial data and uses that to generate the initial structural conditions of opportunities and constraints that characterize the environment of the case study metropolis.

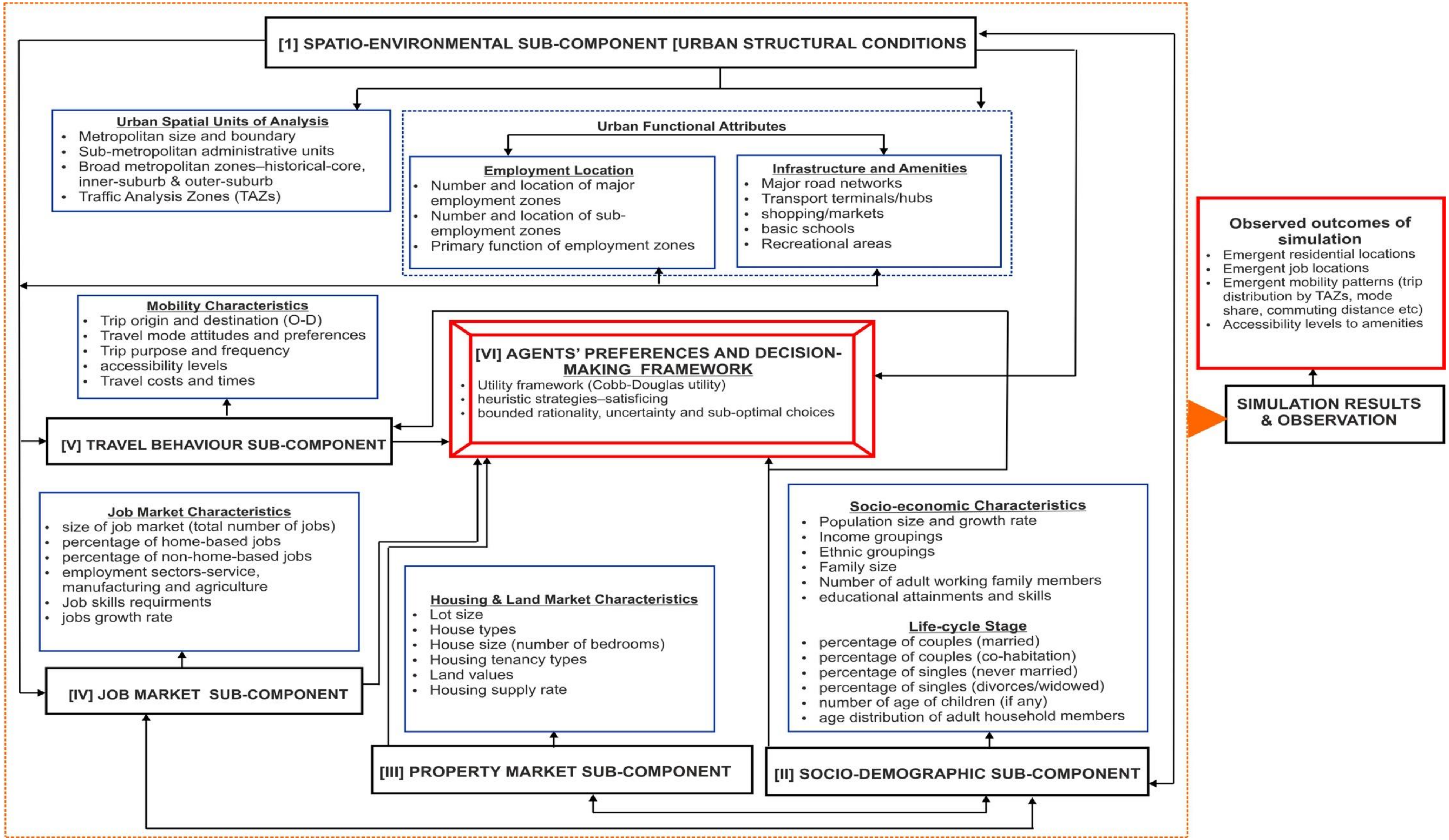


Figure 6.1: Structure of the conceptual model

ii. Socio-demographic sub-component

The socio-demographic sub-component specifies the model entities and their characteristics. In this model, the main decision-making entities are heterogeneous households who make residential decisions and individuals within the households who make job location and travel decisions. A household consists of a family unit and could be an individual or a group of persons who live together and share house-keeping arrangements.

As with existing agent-based models (see for example Tannier et al., 2015; Li and Liu, 2006; Hosseinali et al. 2013; Haase et al.., 2010), households as the primary decision-making agents are differentiated based on socio-economic characteristics (i.e. income, ethnicity, levels of educational attainment, skill levels and number of adult working members), and life-cycle stage attributes such (i.e. marital status, family size, the presences of children and age of adult members). The socio-demographic sub-component also provides the foundation for the encoded mechanism by which new household agents are generated endogenously. A detailed description of the procedure used to initialize and differentiate household agents in the model is presented later in chapter seven where the empirical model is implemented.

iii. Property market sub-component

As depicted in figure 6.1, the property market sub-component generates the housing and land market conditions as the main spatial goods that constitute the object of households' residential choice. It defines the types of dwelling and their intrinsic attributes as well as land parcels, their size and prices. The valuation of a dwelling unit χ^D is a function of its attributes, immediate meso-scale locational attributes and the wider environment within which it is located. This is expressed as follows:

$$\boldsymbol{\chi}^{\mathbf{D}} = \mathbf{f}(\boldsymbol{\chi}^{\mathbf{Ds}}, \boldsymbol{\chi}^{\mathbf{Dt}}, \boldsymbol{\chi}^{\mathbf{Do}}, \boldsymbol{\chi}^{\mathbf{Dp}}, \mathbf{U_Z}, \mathbf{P^x_{amty}}) \qquad \textbf{(6.1)}$$

Where; χ^{Ds} denotes dwelling size (i.e. number of bedrooms);

χ^{Dt} denotes dwelling type (e.g. detached, semi-detached, flat and compound);

χ^{Do} denotes occupancy/tenure type offered with the dwelling;

χ^{Dp} denotes dwelling ask price for rental or owner-occupier tenancy;

P^x_{amty} denotes proximity of dwelling to amenities (e.g. school, shopping, terminals)

U_Z denotes one of three urban-zones (i.e. historical-core, inner-suburb and Outer-suburb) in which the dwelling is located.

The macro and meso-level spatial characteristics combine with the intrinsic attributes of the dwelling to determine the unique characteristics of spatial goods evaluated by households in the residential location choice process.

iii. Job market sub-component

The job market sub-component, as shown in figure 6.1, links with the employment locations defined by the spatio-environmental component of the model to generate the initial job market conditions in the model. Following the assumption of Batty (2005), the main employment zones are exogenously determined and become the active locations where would be individual working members of the household look for employment. Within each employment zones are jobs categorized by work industry and by skills requirements for the job. There exists a direct feedback between the number of active individual agents and the number of available jobs. As will be described later at the model implementation stage, this feedback mechanism mimics a demand-supply interaction by which new jobs are created within the model to match increasing population over time. Moreover, the initial conditions of the job market sub-components provide the basis for home-based jobs—jobs located within the immediate vicinity of the dwelling of agents in the model to emerge endogenously. Details of the characterization and evolution of the job market are provided under the model implementation processes in chapter seven.

iv. Travel choice sub-component

The travel choice sub-component depends on the other three sub-components of the model. it is directly shaped by the urban structural and functional attributes such as the location of jobs and amenities as well as the socio-demographic. This sub-component of the model tracks the travel choices of individuals as well as the emergent patterns of mobility (i.e. trip production and attraction, work-home distances and travel mode choice) that results from the home-work location combinations attainable within the urban context.

6.4.2 Design concepts underpinning the conceptual model

Agent-based models incorporate a range of principles and concepts in their design that distinguish them from other modelling approaches. In the sections that follow, the design concepts and principles underpinning the conceptual model are outlined.

- Sensing, interaction, learning and adaptive behaviour

The purposive household agents represented in the model are receptive to the landscape configuration, and can perceive attributes of the landscape and latent market information with some limitations imposed. For example, household agents can access and randomly sample dwelling units from a sample frame of dwellings in the urban area to assess their attributes such as the type of dwelling, size and prices. They can also perceive, within a limited range, the different combinations of amenity proximity values at selected locations and evaluate them against their preferences to decide their final residential locations. Moreover, household and worker agents can perceive distances from their respective location to key opportunity locations such as major employment areas or local service centres to access available job opportunities.

Furthermore, household agents interact with each other and their environment which leads to changes in the attributes of spatial goods. For example, they engage in competitive market transactions in which they offer bids based on prevailing ask prices and their willingness to pay—bilateral transactions. Based on information made available to them, household agents in the model can learn and modify their behaviour. For example, households adjust bid prices for land or housing to compete favourably in the property market and to benefit from the surplus of trade. The outcomes of the bilateral transaction in turn, determines the formation and evolution land and house prices. Agents of similar socio-economic and life-stage attributes may locate near each other, resulting in some visible patterns of residential clusters within broad urban zones. Also, prevailing job market conditions could lead to individual worker agents settling for home-based employment.

- Feedback

Several dynamic feedback relationships underpin the model design and its implementation at the household level and at the larger urban scale. Feedback relationships exist between population, dwelling units and jobs. These determine the demand and supply dynamics in the model. For example, new dwellings and jobs are supplied in response to population increase over time. New households are formed over time from existing households resulting in endogenous population growth. The location of urban functions also determines the number and attributes of spatial goods, which in turn, shapes location outcomes and mobility patterns such as travel distance, travel costs and mode use.

- Stochasticity

Several stochastic processes underpin the model design. Stochastic processes are used in both the initialization and execution phases of the model. The population sub-model runs on empirically determined probabilities with respect to coupling, child birth and household composition. Age and gender are randomly assigned to populations in the model and calibrated based on observed demographic profiles. The starting locations of the residential-job location choice process as well as the number of choice alternatives are based on empirically determined probabilities. In addition, choice between home-based and non-home-based employment locations as well as work travel mode choice are determined probabilistically.

- Emergence and observation

Several emergent phenomena are observed as model results and outcomes. Local interactions among competing households shape demand and supply dynamics in the property market and ultimately determines the emergent price levels in the housing and land markets. Moreover, total population, population structure and household characteristic in the model are endogenously determined based on an initial sample population. As results of household agents fulfilling their preferences for different urban locations based on their preferences and incomes, residential location patterns emerge. Similarly, employment location patterns emerge from individuals interacting with available jobs within the major employment zones considering home-work distance separation, job availability and skills match between job openings and prospective job seekers. Patterns of spatial interaction in the form of trip origins and destinations (O & D matrix) measuring interaction flows between zones and activities in turn, emerge from home and job locations realized by the agents.

6.4.3 Overview of agents' decision-making framework

The sixth sub-component is at the heart of the overall model framework. It integrates conceptually, all the variables and elements specified into a decision-making framework used by the agents to arrive at their residential location, job location and travel choices. The decision-making framework for each of these choice behaviours are outlined briefly as follows.

- Residential location choice decision-making framework

Purposive household agents use a hybrid utility and heuristics framework to accomplish the task of choosing a suitable location as place of residence. The utility at any location is a function of attributes formalized as follows:

$$\ddot{\mathrm{U}}^{\mathbf{o}}_{(\mathbf{x},\mathbf{y})} = \mathbf{f}(\mathbf{U_Z}, \boldsymbol{\chi}^{\mathbf{D}}, \mathbf{P}^{\mathbf{x}}_{\mathbf{ij}}, \mathbf{P}^{\mathbf{x}}_{\mathbf{amty}}) \tag{6.2}$$

Where:

$\ddot{\mathrm{U}}^{\mathbf{o}}_{(\mathbf{x},\mathbf{y})}$ denotes utility of a location;

$\mathbf{U_Z}$ denotes urban zone of the location (x, y)

$(\boldsymbol{\chi}^{\mathbf{D}})$ denotes the attributes of the dwelling at location (x, y)

$\mathbf{P}^{\mathbf{x}}_{\boldsymbol{ij}}$ denotes distance between employment location (i) and residential location (j)

P^{x}_{amty} denotes proximity of dwelling to amenities (e.g. school, shopping, terminals)

In deciding their place of residence, the household agents use a modified Cobb-Douglas[28] (Douglas, 1928) utility function indicated below:

$$\ddot{\mathrm{U}}^{\mathbf{0}}_{\mathbf{HA}(\mathbf{x},\mathbf{y})} = \prod_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} [\boldsymbol{\mathcal{Z}}\mathbf{i}\ (\mathbf{x},\mathbf{y})]^{\boldsymbol{\alpha}\mathbf{iHA}} \tag{6.3}$$

Where:

$\ddot{\mathrm{U}}^{\mathbf{0}}_{\mathbf{HA}(\mathbf{x},\mathbf{y})}$ denotes the utility of location (x, y) for household HA

αiHA denotes preference weight the household HA places on factor i

$\mathcal{Z}\mathrm{i}(\mathrm{x},\mathrm{y})$ denotes the value of factor i at location (x, y)

n denotes the number of factors evaluated

The utility function assumes that there is a given number of locational factors or opportunities (e.g. shopping, transport terminals, roads and schools) in the metropolitan area for which there are values called proximity index, calculated as the distance between that location and each of the locational factors. Also, for each of the locational factors, a household decides a preference

[28] The Cobb-Douglas utility function is used due to its analytical tractability. The function could take an additive (see e.g. Loibl and Toetzer, 2003; Haase et al., 2010) or multiplicative form (see e.g. Fontaine and Rounsevell, 2010; Brown and Robinson, 2006; Brown et al., 2008). The use of the multiplicative form eliminates the possibility that a location with zero suitability on one factor will have a non-zero utility (Brown et al., 2008).

weight which ranks them in order of importance. The proximity indexes are normalized between 0 and 1 while the preference weights across all the proximity factors evaluated by the household are constrained to the sum of one. Constraining the preference weights to the sum of one normalizes the utility function, achieves easy analytical tractability and allows for the levels of utility of different agents to be compared (Brown et al., 2008).

Households' preferences are constrained by their budgets **(β),** based on their income levels, which is formalized as follows:

$$\boldsymbol{\beta} = (И_{\mathbf{P|r}} + \mathbf{T_{cost}} + \Omega),\ \ \boldsymbol{\beta} = \text{Ï} \tag{6.4}$$

Where:

β	denotes households' budget
$\mathrm{T_{cost}}$	denotes Transport costs between residential locations and employment locations
$И_{\mathrm{P\|r}}$	denotes portion of income on housing—perception of affordability
Ω	denotes a composite good comprising all other non-housing and non-transport spending
Ï	denotes total household income[29]

Using the utility framework and their budgets, households engage in competitive interaction in the property market to fulfil their residential location preferences. The conceptual scheme of market interaction between households and prevailing conditions in the property market resemble initial formulations proposed by Ettema (2011) and Parker and Filatova (2011). Each household allocates a portion of their total disposable income to spend on housing which also reflects their perception of affordability and how much they are willing to pay (WTP) for housing. Household agents make the decision to submit a bid—Willingness to bid (WTB) for either renting or buying a house of some defined characteristics if the ask price of the house is less than or equal to the households' WTP. If a household's WTP is equal to or greater than the ask price of the property, then the household may become the successful bidder subject to the bids of other households who might have also settled on the dwelling as the potential

[29] Total Income refers to the sum of all incomes earned by working members of the household for multiple worker households. It is assumed that for families, both heads of the household earn incomes and that total household income is the sum of individual working member's income. All household expenditures are made from this total income.

property to settle in. Thus, competition ensues between the prospective buyers or renters. The outcome of the competition allocates the property to the highest bidder. In the event of a tie, one of the households is randomly allocated the property. This approach allows to endogenously model expectation-driven price dynamics resulting from interaction among buyers and sellers, to realistically represent real-world housing market dynamics.

The use of the utility framework as decision-making criteria does not presuppose that households only arrive at optimal locations. Instead, uncertainty and bounded rationality are introduced, which can result in sub-optimal outcomes to the location choice process. These are integrated in a heuristic strategy used by the household agents. Firstly, households do not have information about the entire characteristics of the property market. Rather, the amount of information is limited to a given search radius which samples a limited number of location options to evaluate. Moreover, agents use *satisficing*[30] strategy as a heuristic technique (Leong and Hensher, 2012). By satisficing technique, household agents do not necessarily have to exhaust the entire list of selected options within the search radius before deciding. Instead, they stop at the alternative that meets their preferences regardless of the number of alternatives that have been evaluated previously from the pool of randomly selected dwelling units.

A detailed description of the residential location search process, the schedule of the search tasks and the heuristics implemented by the household agents is presented during the implementation in Chapter seven.

- Job location choice decision-making framework

Job location choice constitutes the second important task performed by agents in the model. Whereas residential location decisions are made by the household as a family unit, when it comes to job location choice, only the individual adult members in the family are the actors. Job location outcomes is a function of:

$$\mathbf{J_l = f(E_n,\ E_{jv}, P_{ij}^x,\ S_m\ )} \qquad (6.5)$$

Where;

[30] The adoption of satisficing technique in the initial model is intended reduce model computation demands whilst mimicking the real world-decision-making behaviour of household agents faced with uncertainty and limited access to information. It is possible to implement other heuristic strategies, and these will be explored during the model implementation.

J_l denotes job location of an individual worker within a household;

E_n denotes number and location of employment opportunities;

E_{jv} denotes the availability of vacant jobs at the employment locations

P_{ij}^{x} denotes the distance separation between potential employment and residential locations; and

S_m denote the skills match between the available vacant jobs and the job seeker

Job seekers within each household can perceive the location of the major employment locations as well as the home-work distance separation. They also receive information regarding job vacancy and the skills requirement for jobs at any of the employment zone. Worker agents seek employment at one of the employment zones, weigh vacancies based on skills match, and decide on positions. Where there is more than one job seeker for a single job opening, with all jobs seekers possessing skills that match the available job, the job is offered randomly to one of them.

Besides seeking employment within one of the employment zones, there is the possibility for home-based jobs to emerge. Thus, some workers, depending on the context, could decide to engage in home-based self-employment. The decision to engage in home-based employment as will be discussed later in chapter seven, is modelled endogenously as a function of the skills and educational attainment of the job seeker, the urban-zone of location (i.e. whether the job seeker resides in the historical-core, inner-suburb or outer-suburb) as well as the initial job search outcomes.

- Commuting preferences decision-making framework

In addition to making employment location decisions, individuals make choice decisions regarding their home-work commuting preferences as shown in equation 6.9.

$$\mathbf{C_{prf} = f(\ddot{I}\,,\ U_Z\ ,\ j_{\mathit{loc}}\ ,D_{ij}\ ,T_{mode}, T_{cost})} \qquad (6.6)$$

Where;

C_{prf} denotes commuting preferences of an individual worker;

$\ddot{I}$ denotes income group of the worker;

U_Z denotes urban zone of residence (i.e. historical core, inner-suburb, outer-suburb);

J_{loc} denotes job location of individual worker (i.e. home-based and non-home-based)

D_{ij} denotes distance between the place of residence i and the place of work j

T_{mode} denotes transport mode options available in the urban area

T_{cost} denotes transport costs between residential locations and employment locations

Thus, the above elements of individuals' commuting preferences are shaped by their residential and job location combinations, which in turn, determine important daily travel decisions regarding transport mode choice and the associated travel costs.

6.5 Chapter summary

In this chapter, a methodological framework to simulate urban location choice and mobility patterns has been specified. In line with current research trends and the goal of this research to develop a disaggregate model using a novel theory grounded technique, ABM paradigm was adopted for the conceptual model specification. The ABM approach derives strength from making it possible to represent, in a spatially explicit manner, the micro-level attributes that shape the location and travel choice behaviour of households and individuals.

In line with the purpose of the model, six main sub-components, each specifying the key entities and variables required, were identified in the conceptual model. These were the spatio-environmental, socio-demographic, property market, job market and travel behaviour sub-component and agents' behaviour and decision-making framework. Relevant ABM design concepts underpinning the conceptual model including sensing, learning, adaptive behaviour, interaction, feedback and stochasticity were also discussed.

A decision-making framework, combining a utility framework with principles of bounded rationality and heuristic strategies was specified to guide households' residential location choice behaviour. For job location choice, a framework comprising the spatial distribution of jobs, characteristics of the jobs (e.g. skill requirement, vacancy status) and characteristics of the potential job-seekers was specified. Finally, individuals travel decision was formalized as a function of the interplay between a set of socio-economic and spatial factors including income, urban zone of residence, home-work distance separation, travel mode options available and the associated transport costs.

The rationale of the model framework presented in this chapter was to specify a methodological approach that would provide the relevant theoretical background, modelling paradigm and behavioural framework as the basis for implementing an empirical model. The next chapter addresses the model implementation process. It provides a detailed description of how the conceptual model specified in this chapter has been applied and programed within a computer simulation environment to simulate the urban location choice and mobility nexus of the case study metropolis.

CHAPTER SEVEN: IMPLEMENTATION OF AN INTEGRATED GEOSPATIAL AND AGENT-BASED MODEL OF URBAN LOCATION CHOICE AND MOBILITY PATTERNS

7.1 Introduction

This chapter details the programming of the integrated geo-spatial and agent-based model to simulate the co-emergence of urban location choice and mobility patterns, which was indicated as the second objective of this thesis. Utilizing the conceptual model presented in chapter six and interpreting the results of the empirical data analysis presented in chapters four and five, this chapter formulates the decision-making rules governing agents' behaviour, the scheduling of tasks implemented by agents, and the rules and the heuristics used by the agents to accomplish their location and travel choice objectives. It also discusses how these processes and procedures have been encoded within a computer-based simulation software using pseudo codes represented in flow diagrams of simplified condition-action-rules.

The development of the model is grounded in both the theoretical proposition and empirical evidence of the relationship between location decision outcomes and patterns of spatial interaction in urban areas. Decisions taken by households and individuals with regards to where to live and where to work are complex in nature. This is because from the behavioural point of view, these decisions are taken by actors who exhibit heterogeneity in socio-economic status, socio-demographic characteristics, perceptions and preferences. The urban landscape within which the diverse actors make their location choice also exhibits heterogeneity at different spatial scales and over different time horizons. Moreover, these decisions have long-term consequences: the outcomes of the random location decisions of individuals and households collectively, impact urban spatial structure, which in turn, shape daily mobility characteristics, including trip production and attraction patterns and travel mode choice. The ability to model residential and job location decisions therefore offers great potential to understand and to predict how urban spatial structure co-evolves with patterns of spatial interaction.

The model presented in this chapter, hereafter referred to as the Metropolitan Location and Mobility Patterns Simulator (METLOMP-SIM), has been developed to simulate the residential location choice of households, the job location choice of individuals within the households and the accompanying home-work mobility choices, using the Kumasi metropolis in Ghana as case study. As a facsimile model, METLOMP-SIM is the culmination of the two empirical studies from the case study area on urban location choice and mobility characteristics presented in chapters four and five. The model is spatially explicit, capturing the unique attributes of location and spatial goods at multiple spatial scales. Moreover, the dynamics of the urban property markets involving the bilateral transactions and competition among agents underpin price formation and evolution in the model.

The implementation of METLOMP-SIM is driven by the following research questions:

i. How do the socio-demographic characteristics and preferences of heterogonous households and individuals interact with existing urban structural conditions to influence urban location choice behaviour?
ii. How do bilateral transactions, competitive behaviour and interactions among individual actors in the property market lead to the formation and evolution of property prices?
iii. What are the residential location patterns that emerge from the interaction between households 'and individuals' choice behaviour and existing urban structural conditions?
iv. What are the employment location patterns that emerge from the interaction between the attributes of individual working members of the households and prevailing job market conditions?
v. How does the emergent residential and job location combinations and individual-level attributes of agents interact to shape home-work mobility patterns?

7.2 Chapter organization

The rest of this chapter is organized under six main sections. The model implementation begins with a brief discussion of existing agent-based modelling platforms, justifying the simulation platform and programming language adopted for the development of METLOMP-SIM. Next, an overview of the model's input data is presented. This is followed with a detailed discussion of the encoded procedures used to generate the initial conditions of the model under each of METLOMP-SIM's sub-components. In the fourth section, the decision-making tasks of the agents, the programmed condition-action-rules and heuristics used by the agents to perform these tasks, as well as the sequential scheduling of the tasks are outlined and discussed. In the concluding section, a summary of the model implementation process is presented to pave way for the discussion of the model calibration and simulation results in the next chapter.

7.3 Model implementation platform and programming language

The implementation of METLOMP-SIM involved translating the conceptual model into computer codes and algorithms. Since developing a software platform from the scratch requires domain expertise as well as time, an off the shelf, bespoke ABM software was considered suitable for the development of the model.

Many computer software programs exist for developing ABMs. Each of these platforms differ in their capabilities and in the modelling toolkits they offer, and reflect the design objectives and philosophies of the developers (Railsback et al., 2006). In choosing the right software to implement METLOMP-SIM, authoritative reviews on existing ABM platforms were first consulted (see e.g. Gilbert, 2008; Gilbert and Bankes, 2002; Railsback et al., 2006; Tobias and Hofmann, 2004). These platforms were evaluated against the purpose of the model to be developed, the functionalities required and user-friendliness to both the researcher and others who would interact with the final model.

Among the existing ABM platforms such as Swarm, MASON and Repast, Netlogo[31] (Wilensky, 1999) emerged as the most suitable software for the implementation of METLOMP-SIM. Netlogo is a widely-used software for developing ABMs. The platform provides a simple yet powerful Object-oriented programming language, built-in graphical interfaces and comprehensive documentation, all of which makes it relatively easier to use. The software also allows as much flexibility as possible to represent agents as heterogeneous decision makers while providing GIS functionality to capture various attributes of the urban space as required in METLOMP-SIM.

METLOMP-SIM is implemented in Netlogo version 5.3. Although the model is coded in Logo, Netlogo's programming language, the exact computer codes and algorithms are not reproduced

---

[31] Netlogo is a free software platform developed at the Centre for Connected Learning and Computer-Based Modelling (CCL), Northwestern University, Evanston, Illinois. The version 5.3 used in this research was downloaded at: *https://ccl.northwestern.edu/netlogo/download.shtml*

here. Instead, as will be discovered later in this chapter, pseudo codes in the form of flow charts and simple condition-action statements are used to illustrate the encoded procedures in the model.

7.4 Overview of model input data

The model's input data consists of two main types. The first datasets comprise geographically referenced spatial datasets on the Kumasi metropolis, processed using GIS software. An outline of the input spatial datasets used is provided in table 7.1. Detailed explanation of the data handling processes and how each layer of spatial data fed into the sub-components of the model are discussed later under model initialization in section 7.5.

Table 7.1: Spatial datasets for model implementation

Datasets	Description
Administrative boundaries data	GIS shapefile, metropolitan and sub-metropolitan boundaries
Location of Infrastructure and amenities	GIS data of road hierarchy: arterials, distributors and access roads, transport terminals, schools at metropolitan, sub-metropolitan and settlement scales
Location of dwelling units	GIS shapefile of Metropolitan, sub-metropolitan and settlement scales
Location of employment zones	GIS shapefile of major employment zones located in the Kumasi metropolis
Metropolitan TAZ system	Maps of Traffic Analysis Zone system obtained and digitized in GIS
Historical land values	Land transactions data from 2000, 2005, 2010 and 2015, translated into GIS raster file using interpolation function
Location of natural urban features	GIS shape file of rivers, forests at the metropolitan, sub-metropolitan and settlement scales
Spatial buffers	GIS shape file of multiple buffers around the metropolitan centre and along major road networks to track land-price evolution.

Aspatial information from the analysis of primary data obtained through the cross-section survey of households and individuals in the Kumasi metropolis constituted the second set of input data for the model. Figure 7.1 illustrates conceptually, the linkages between the survey data analysis results presented in chapters four and five and the model sub-components. The linkages between each of the survey data analysis themes and the model sub-components are depicted using arrows directed to the latter from the former. For example, results of the analysis of the background socio-demographic attributes of households and preferences for residential location factors presented in chapter four, feeds into the socio-demographic sub-component and agents' preferences and decision-making rules of the model.

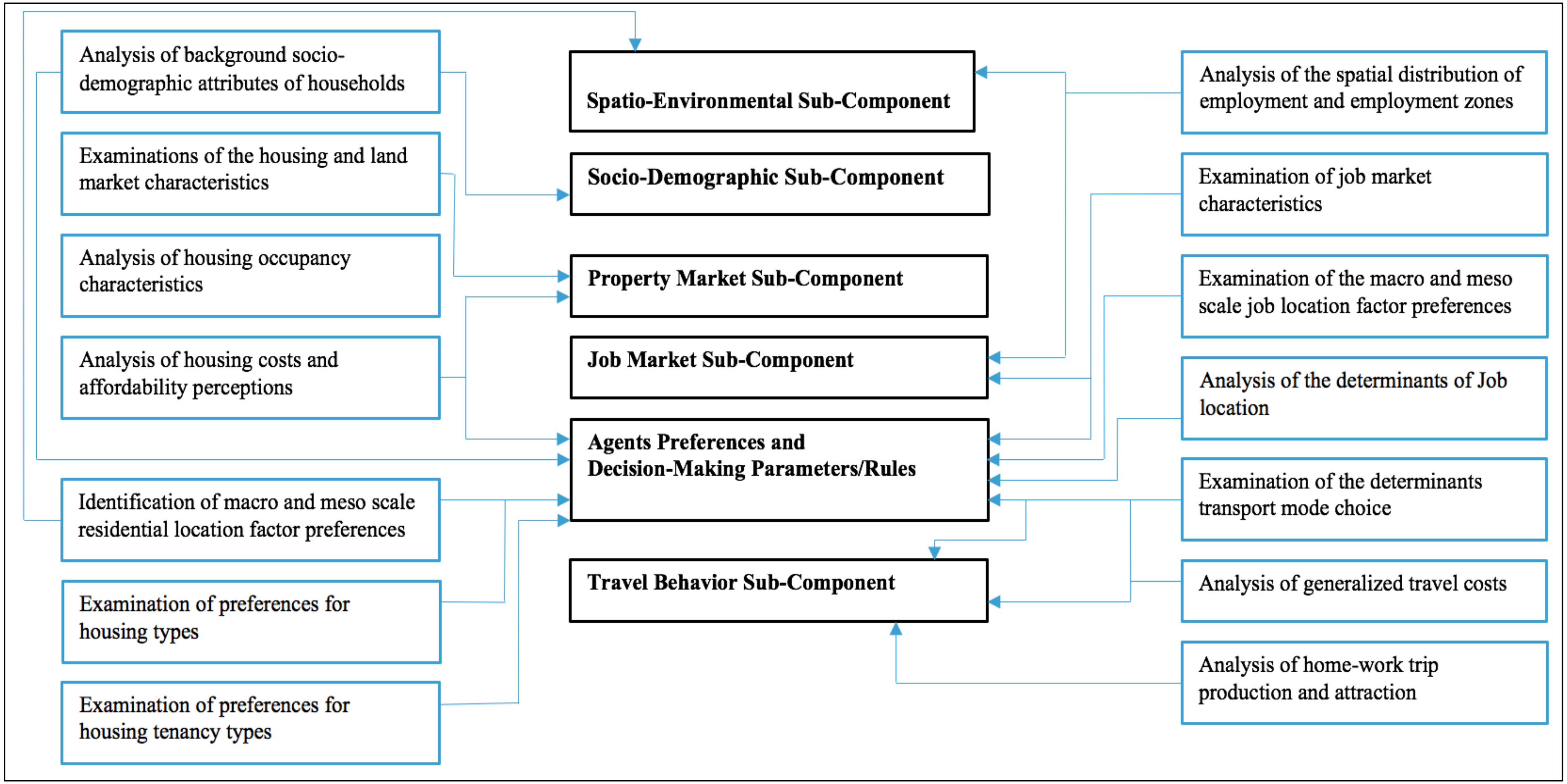


Figure 7.1: Conceptual illustration of the linkages between the model sub-components and survey data analysis themes
NB: Model sub-components are presented in black outlined box in the middle while the input results of the survey data analysis, organized around themes 14 themes are presented in blue outlined box on the left and right hand side

Exactly how the results of the survey data analysis have informed the model implementation will be highlighted later under the relevant sections addressing the initialization (i.e. section 7.5) and execution (i.e. section 7.6) of the model. Notwithstanding, it is worth clarifying that the empirical insights derived from the results of the survey data analysis are not repeated here but are referred to and interpreted for the model implementation.

7.5 Initialization of METLOMP-SIM: Description of model sub-component implementation processes and procedures

The model initialization phase implements the procedures[32] programmed to generate and set-up attributes of the metropolitan environment which constitute the initial conditions required for the simulation. An overview of the model initialization process is presented in figure 7.2.

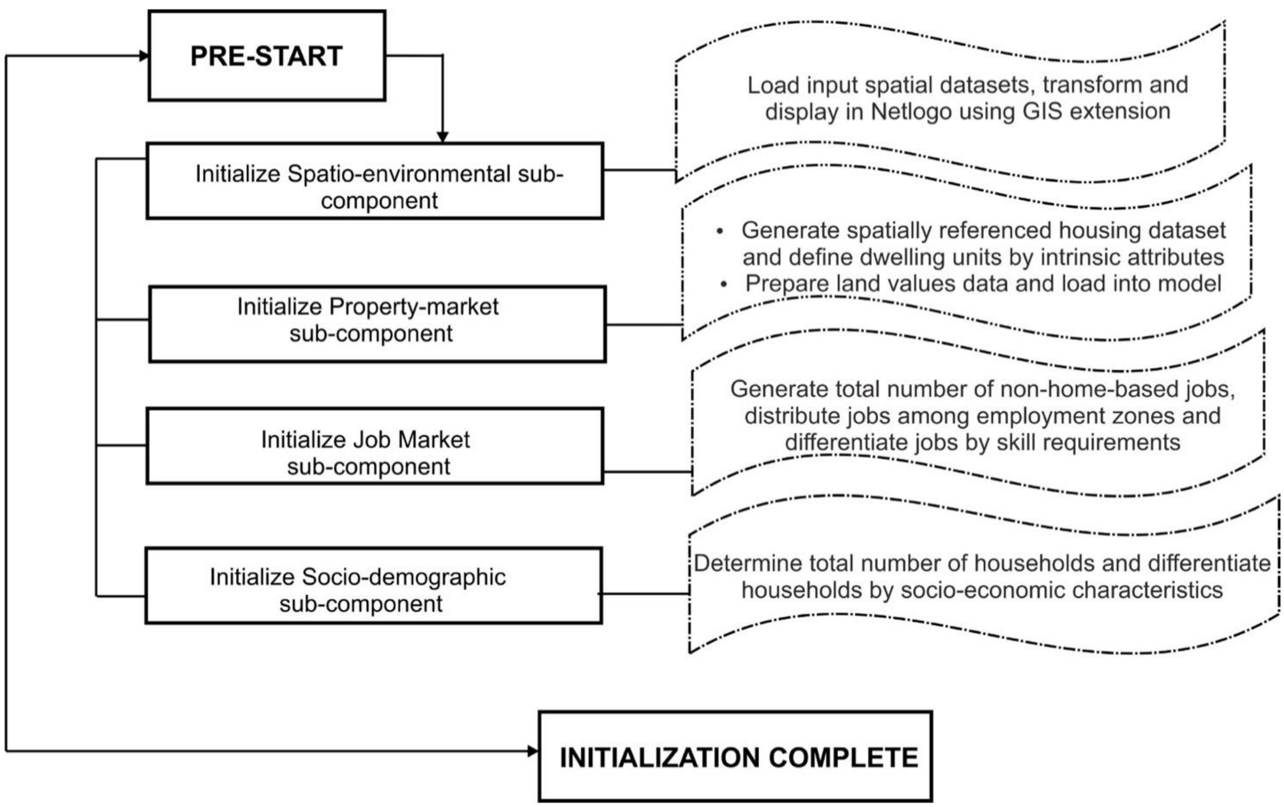


Figure 7.2: Overview of METLOMP-SIM's initialization processes

[32] In Netlogo programming, a procedure is a sequence of commands written to accomplish a specific task

In sequential order, five main processes linked to the model sub-components initialize METLOMP-SIM. This comprises the 'pre-start' procedure, followed by the initialization of the spatio-environmental, property market, job market and socio-demographic sub-components respectively. In the sections that follow, the procedures implemented under each stage of initialization are discussed.

7.5.1 The 'Pre-start' procedure

Prior to initializing the model sub-components, the 'Pre-start' procedure, implements a series of commands which essentially, resets the world/display monitor to an initial, empty state and clears all variables and results of previous simulations. The procedure also implements the 'reset-ticks' command, which sets the tick counter used to monitor time to zero. The model interface generated by the pre-start procedure is shown in figure 7.3.

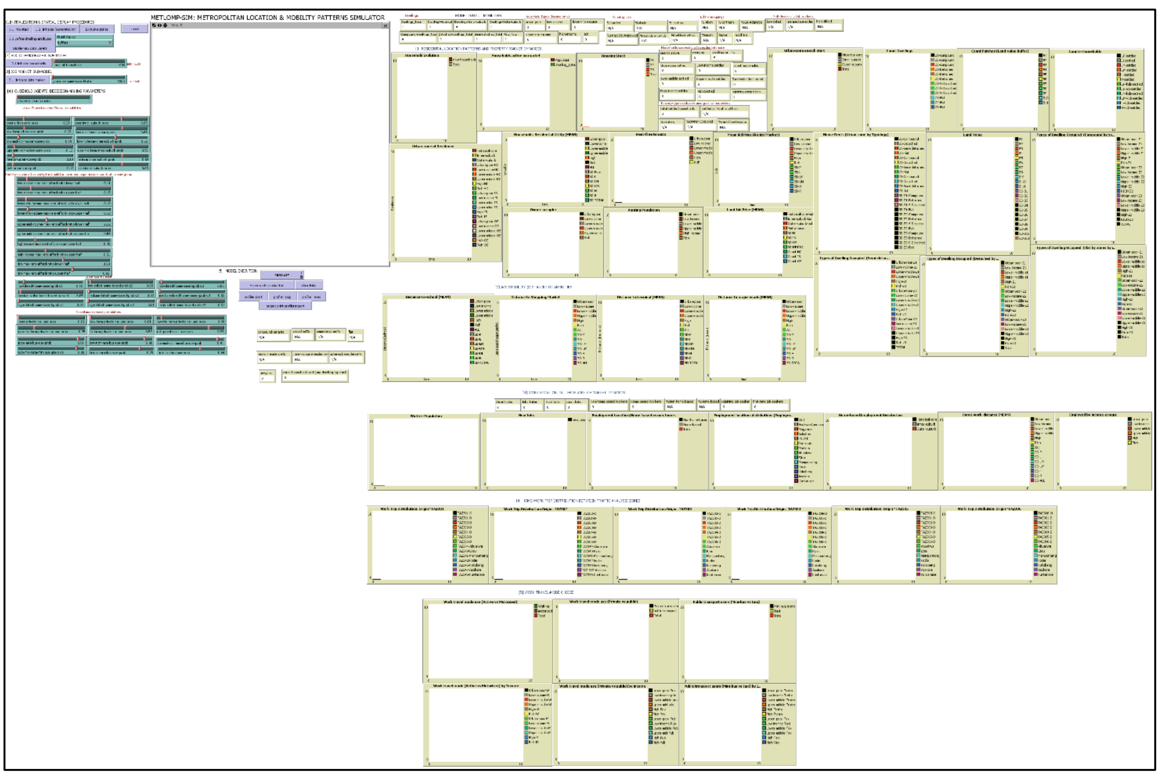


Figure 7.3: Model interface prior to initializing generated by pre-start procedure

7.5.2 Procedure to initialize spatio-environmental sub-component

Using Netlogo's GIS extension, a series of commands are programmed to initialize the spatio-environmental sub-component. The procedure generates the geographical setting of the simulation as well as the various attributes that characterize the metropolitan landscape. The main steps involved in initializing the spatio-environmental sub-component are described as follows:

i. Boundary and spatial features dataset preparation, importation and transformation in Netlogo

The relevant spatial data were first prepared using ArcGIS software. Table 7.2 shows a summary of the spatially referenced data imported into the model when the spatio-environmental sub-component is initialized. All GIS data were projected to WGS 1984 UTM Zone 30N Coordinate System and processed as vector data in shapefile format.

Table 7.2: GIS datasets imported using the spatio-environmental sub-component initializing procedure

GIS datasets	Type	Description of layer
Metropolitan boundary	Polygon	Vector data of the administrative boundary of KMA covering a total area of 212km^2
Sub-metropolitan units' boundary	Polygon	Vector data of the 9 sub-metropolitan divisions within the KMA.
Urban-zone divisions	Polygon	Vector data of three broad urban-zones defined within the KMA—Historical core (22km^2) Inner-suburb (38.7 km^2) and Outer-suburb (145km^2)
TAZ System	Polygon	Vector data of the 29-internal micro and 6 macro TAZs and 6 of the KMA
Primary road-network	Polyline	Vector data showing the major road system in the KMA
Location of transport terminals	Point	Vector data showing the spatial distribution of transport terminals in the KMA
Location of primary schools	Point	Vector data showing the spatial distribution of primary schools in the KMA
Location of local shopping/market centres	Point	Vector data showing the spatial distribution of major shopping/market centres in the KMA
Housing/dwelling units	Point	Vector data showing the spatial location of sampled housing/dwelling units in the metropolis.
Major employment zones	Polygon	Vector data showing the spatial locations of the main employment zones within the metropolis
Nature-reserves and no-go areas	Polygon	Vector data of restricted development areas extracted from the metropolitan land use map.

Using Netlogo's inbuilt GIS data loading commands, each layer of data was imported into the model. The Netlogo default environment is made up of grid cells. This means that, a transformation should be defined between the GIS data and the Netlogo space to merge the two. To achieve this, the inbuilt command which allows to take the union of the "envelopes" or bounding rectangles of the datasets in GIS space and map them directly to the bounds of the NetLogo world was implemented in the second step. At the third stage, the inbuilt *'intersect'* command was used to embed the GIS data into the Netlogo space at the appropriate resolutions. To make the GIS data active for subsequent operations such as querying and addition of attributes, the GIS vector data were transformed into active layers perceptible to agents using *'gis:feature-list-of'* command. At the final stage of the procedure, the inbuilt *'gis:draw'* command was used to draw and display the data in the model interface as shown in figure 7.4

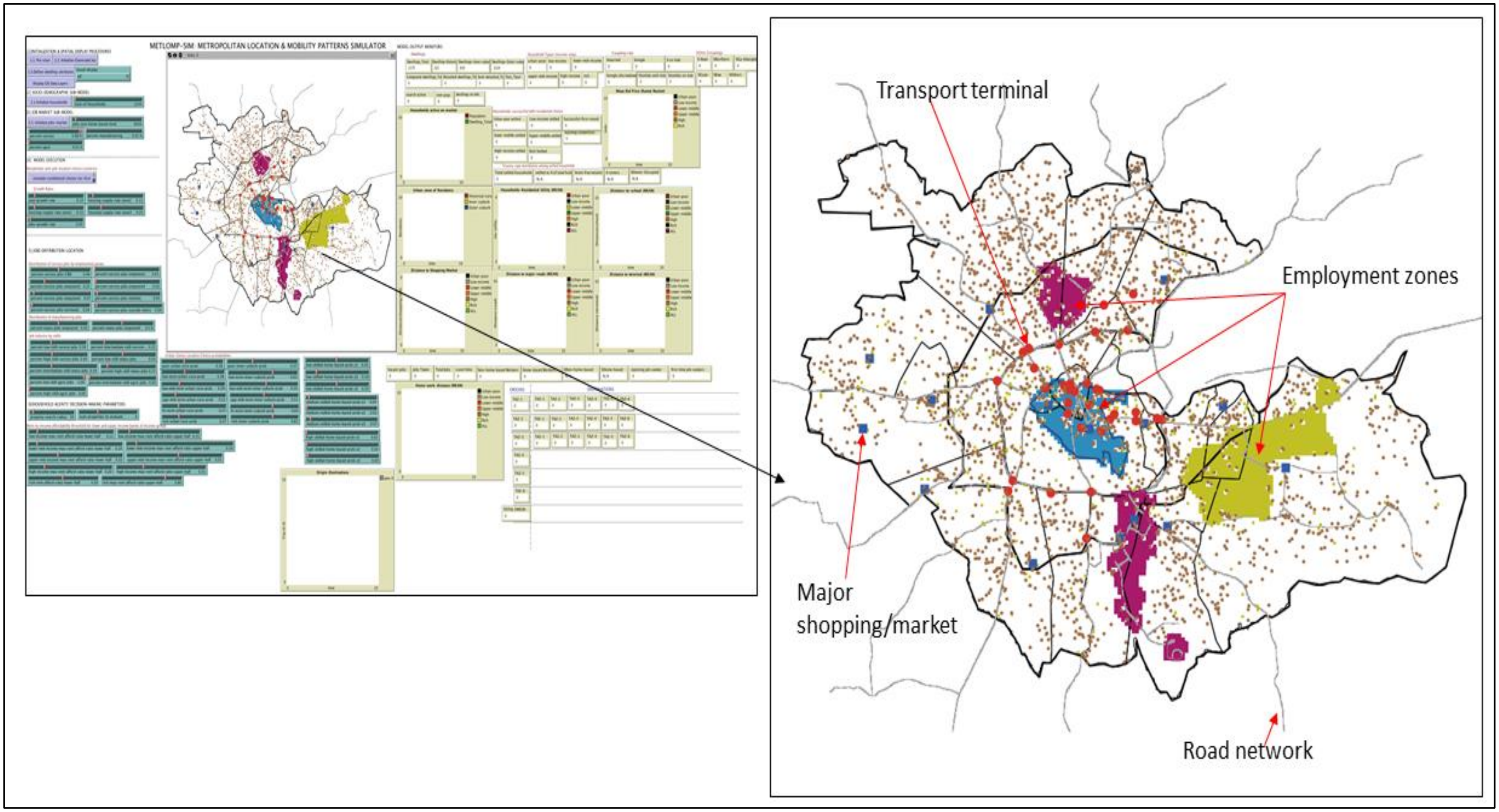


Figure 7.4: Screen capture of GIS data imported and transformed in Netlogo using GIS extension

ii. Computing and loading infrastructure/amenities proximity indexes

As formalized in the conceptual model in chapter six, households use a modified version of Cobb-Douglas utility function (see equation 6.4) in their residential location choice. The main parameters of the utility function are a set of locational factors and their numerical values. The principal component analysis (PCA) of residential location choice factors in the Kumasi metropolis, presented in chapter four, distilled the determinants of location choice into four main components. Two of these latent components extracted by the PCA were proximity to major infrastructure and amenities and proximity to household members core activities (i.e. school of children and work-place of adults). Based on these findings, spatial proximity indices for the factors identified were computed at a 30.48m × 30.48m (or 100ft × 100ft) resolution and imported into the model to characterize the spatial environment of the simulation. The steps involved are summarized in figure 7.5. The proximity indexes were normalized between 0 and 1, loaded into the model and colour-scaled as shown in figure 7.6

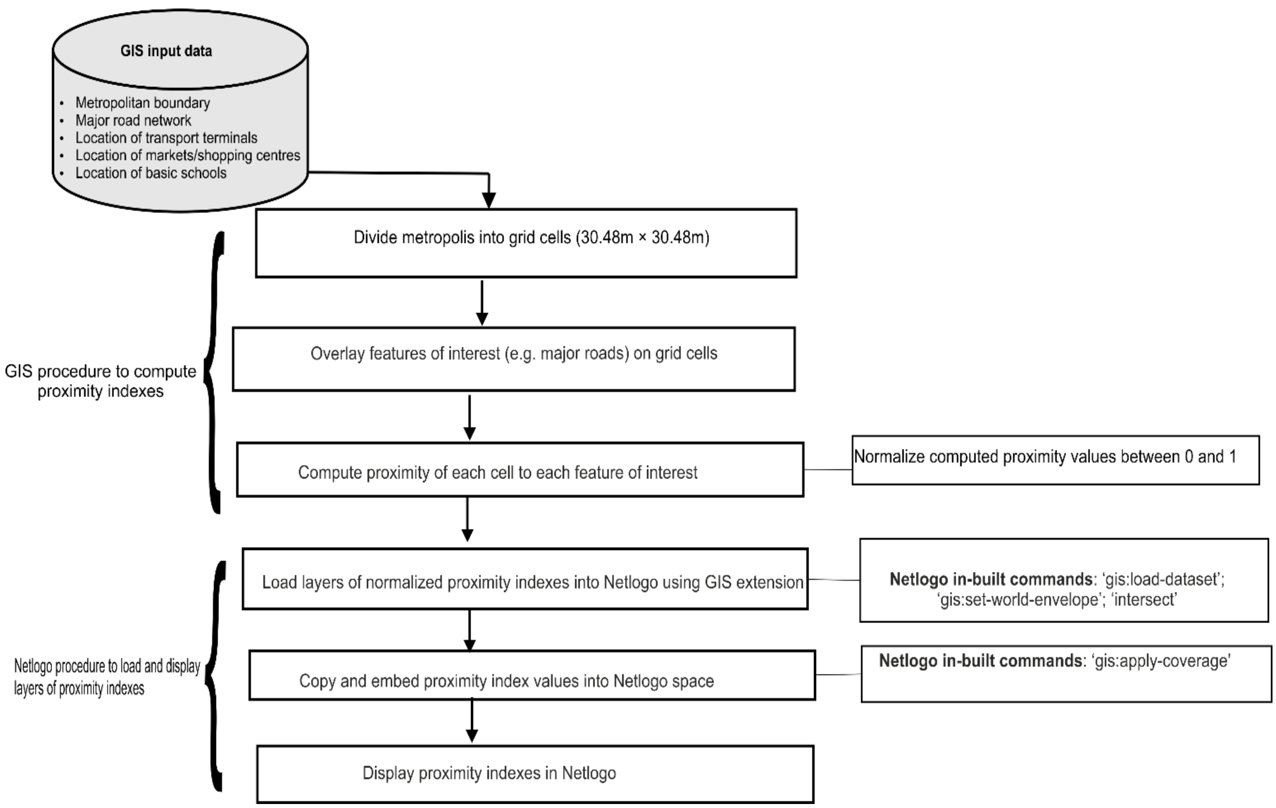


Figure 7.5: Computing amenity proximity indexes in GIS and loading output in Netlogo

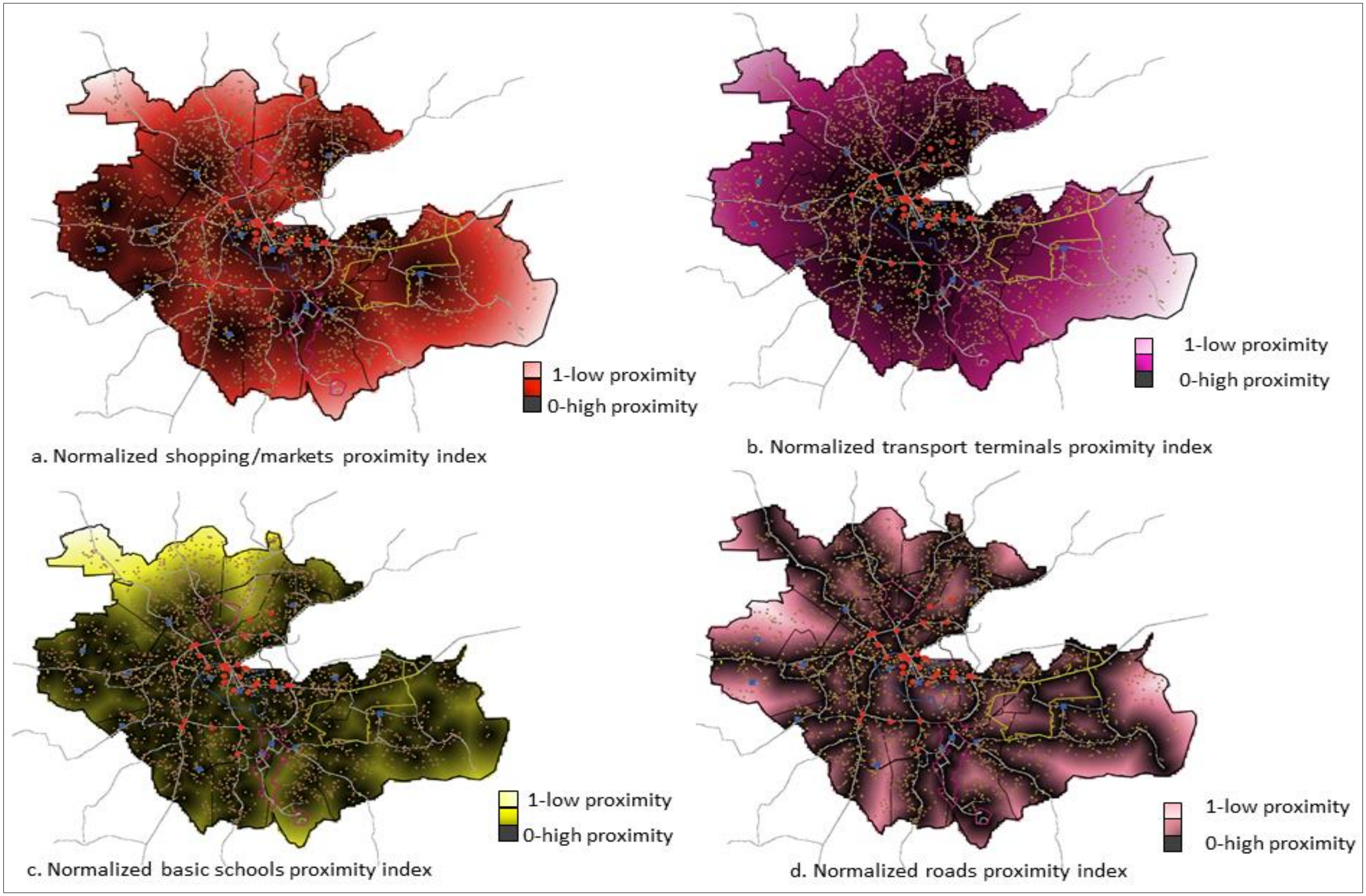


Figure 7.6: Infrastructure/Amenity proximity map layers imported transformed and visualized in Netlogo using GIS extension

7.5.3 Procedure to initialize property market sub-component

The initialization of the property market sub-component implements a series of commands that generate the initial conditions characterising the housing and land markets in METLOMP-SIM. The first procedure implements a set of commands that generate and differentiate the total housing stock in the model. The second procedure involves the preparation of land price map in GIS and importing the output into the model. Each of the procedures are explained in the sections that follow.

i. Generating housing stock and defining dwelling attributes

As specified in the model development framework, dwelling units constitute one of the bundle of choice alternatives in METLOMP-SIM. Spatially referenced housing data of any given sample size can be imported to populate the model. The generation of the housing stock in the model involved six main steps depicted in figure 7.7.

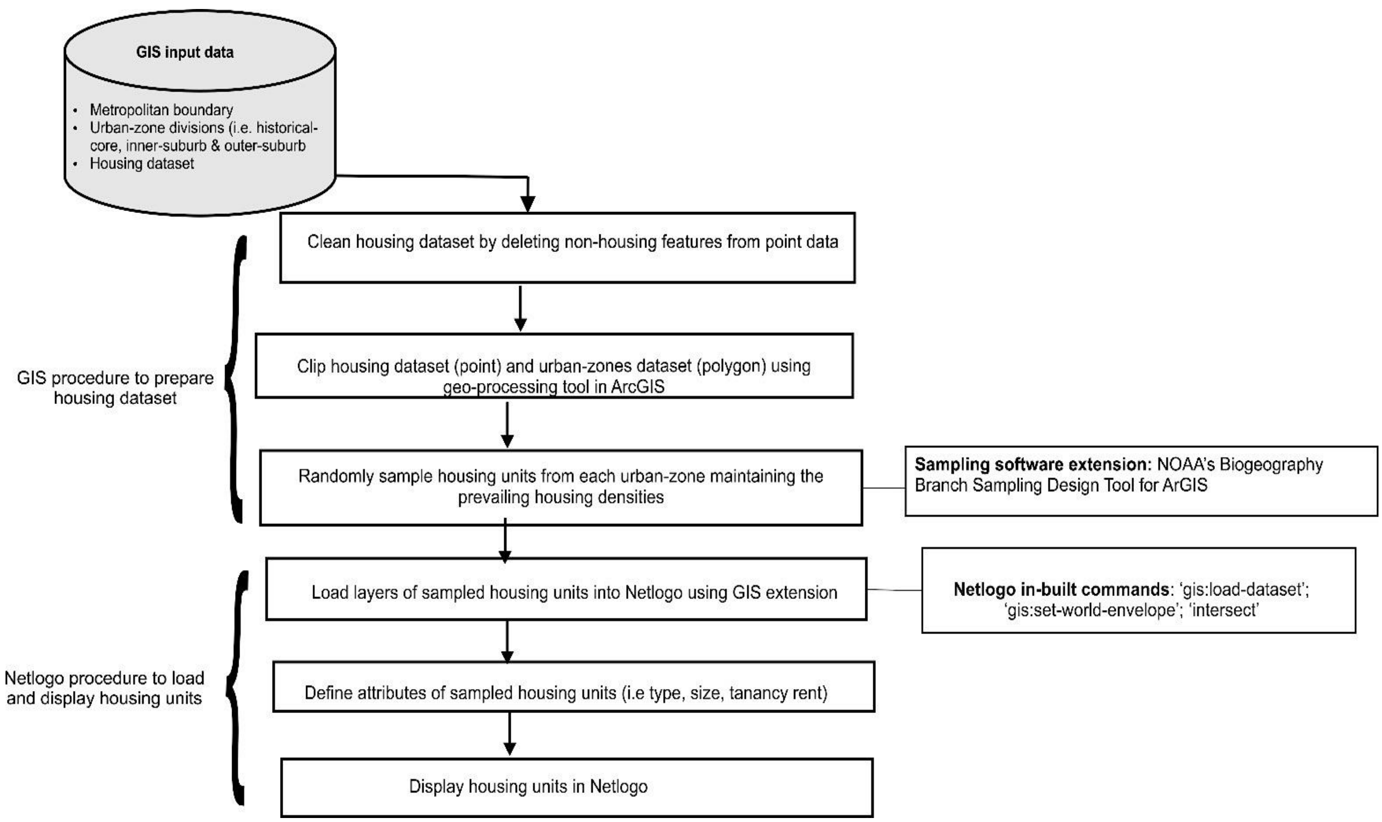


Figure 7.7: Procedure to prepare housing data and load into Netlogo

Where only a sample of the dwellings are needed as was the case of the current version of the model, a random sampling tool— NOAA's Biogeography Branch Sampling Design Tool for ArcGIS[33] was used to select dwellings from each of the urban zones (strata) to be represented in the model for simulation. This sampling tool allows to perform a spatial sampling from a sample frame of point data while maintaining the same density distribution in the sample as the original sample frame. A sample of the housing datasets generated in the model using this tool is shown in figure 7.8.

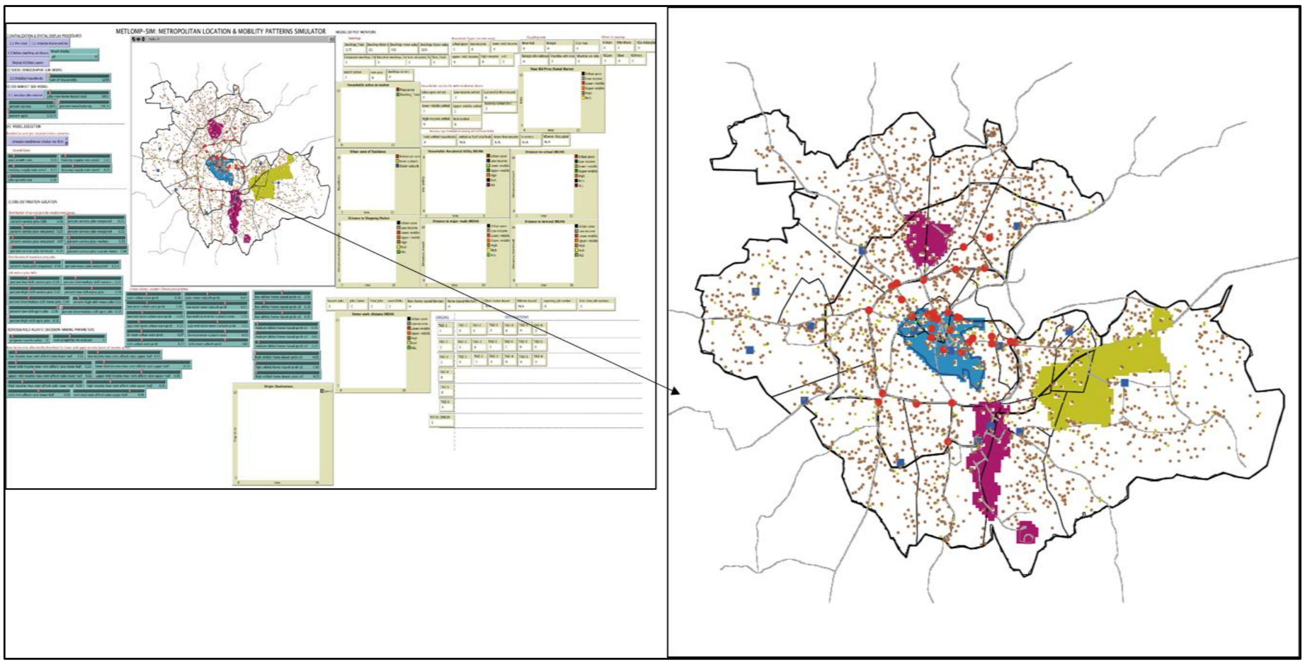


Figure 7.8: Sampled housing stock loaded into Netlogo using GIS Extension. NB: brown dots represents the housing units; in this figure 2% of the total housing stock in each zone have been loaded.

The dwelling units in the model are differentiated by—type, dwelling size, occupancy/tenure type and price/rent. Since the available housing data did not have these attributes, results of the housing market analysis informed the characterization of the housing stock in the model[34]. Dwelling units were first randomly, differentiated into four main types (i.e. compound, detached, semi-detached

---

33 NOAA's Biogeography Branch Sampling Design Tool for ArcGIS is an opens source add in or extensions that is installed in ArcGIS for purposes of spatial sampling. The accompanying instructional manual was downloaded athttp://www2.coastalscience.noaa.gov/publications/detail.aspx?resource=mgAPjEXoVKEOwI+9IXOjBThd9nQ55Og19XX2UqnBEdY= ; accessed, 18 November, 2015)

34 Detailed analysis of the market characteristics discussing the types, size, tenancy arrangements and house rents in the study area was presented in chapter four, section 4.5.

and flat) based on the observed proportions of each of the dwelling types in the three urban-zones. For example, since from the available data, 57% of all dwellings in the historical-core of the case study metropolis was compound, the same proportion is assigned in the model at initialization. Other attributes defining dwellings within the three urban zones such as tenancy type (i.e. owner-occupied, renting and rent-free) dwelling size and rent-band are also specified using a random selection and allocation command in Netlogo based on representations of these characteristics in the observation data.

Moreover, housing rent-bands in the model range between Band-A and Band-J, based on the percentile distribution of rents charged for rental properties in the Kumasi metropolis presented in chapter four, section 4.7.4. For example, based on the findings of the empirical analysis, 67% of 4-bedroom detached dwellings, located in the historical-core of the metropolis, and available for renting was assigned the rent band "I", which translates into a monthly rent amount range of GH₵400 and GH₵500. Setting the initial rent-bands of dwellings involved using the *'random-float'* command which takes the lower and upper limits of monthly rent-band and assigns them randomly to selected dwellings at specified intervals of increment. Table 7.3 provides a summary of attributes defining dwelling units and their proportion representations on initializing the property market sub-component of METLOMP-SIM.

Table 7.3: The distribution and characteristics of dwellings at initialization

Themes	**Variables**	**Values**
Dwelling units	Total dwelling units	Specify at initialization
Urban-zone distribution of dwellings (%)	Historical-core zone dwelling units	15
	Inner-suburban zone dwelling units	29
	Outer-suburban zone dwelling units	56
Dwelling types (%)	Compound dwelling units	46
	Detached dwelling units	25
	Semi-detached dwelling units	15
	Flat (multi-storey) dwelling units	14
Tenancy distribution (%)	Dwelling units available for rent-free	39
	Dwelling units available for renting	42
	Dwelling units available for ownership	19
Detached dwellings size (%)	Detached 1 bedroom	18
	Detached = 3 bedrooms	16
	Detached = 4 bedrooms	19
	Detached = 5 bedrooms	20
	Detached >5 bedrooms	21

Source: Based on Field Survey, February 2015

Table 7.3 continued: The distribution and characteristics of dwellings at initialization

Themes	Variables	Values
Semi-detached dwellings size (%)	Semi-detached = 1 bedroom	26
	Semi-detached = 2 bedrooms	37
	Semi-detached = 3 bedrooms	13
	Semi-detached = 4 bedrooms	13
	Semi-detached = 5 bedrooms	10
	Semi-detached > 5 bedrooms	1
Flat size (%)	Flat = 1 bedroom	35
	Flat = 2 bedrooms	26
	Flat = 3 bedrooms	26
	Flat = 4 bedrooms	2
	Flat = 5 bedrooms	3
	Flat > 5 bedrooms	8
Compound housing size (%)	Compound housing < 5 rooms	11
	Compound housing >5 and < 10 rooms	38
	Compound housing >10 and < 15 rooms	22
	Compound housing >15 rooms and < 20 rooms	17
	Compound housing > 20 rooms	12

Source: Based on Field Survey, February 2015

ii. Interpolation of land values dataset using ordinary exponential semi-variogram method

Since in the case study metropolis, would-be owner-occupiers make land acquisition decisions to develop their housing incrementally instead of buying their homes, land price data in GIS format was required in the model. Land price data dating back to 2000 obtained from 30 locations in the metropolis where data was readily available served as input into the model.

Figure 7.9 shows the distribution of land values in the 30 proxy locations in the Kumasi metropolis. It shows that land prices are highest at the Adum-kejetia area, which also functions as the CBD of the metropolis. In general, land prices tend to fall as distance increases farther away from the core areas of the metropolis. The relationship between land values and distance from CBD, depicted in figure 7.10 shows that for every 1km increase in distance away from the CBD of the metropolis, land values decrease by GH₵26960, (R-squared = 0.1974; P = 0.014).

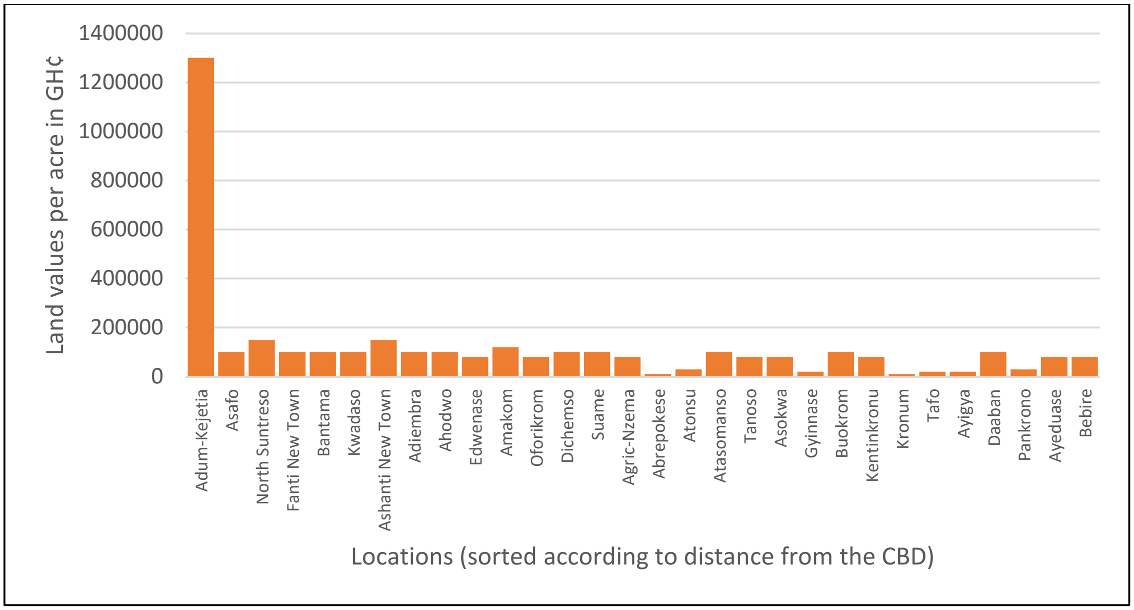


Figure 7.9: Land price data for 2000-base year:
Source: Based on data obtained from Lands Commission, Kumasi

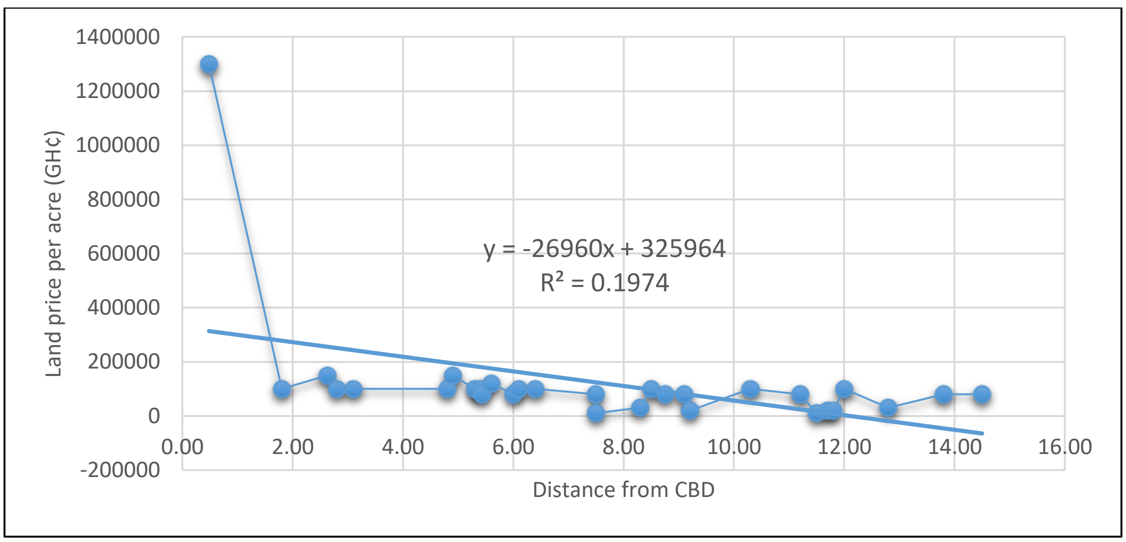


Figure 7.10: Relationship between land values and distance from CBD
Source: Based on data obtained from Lands Commission, Kumasi

In order to derive a map layer of land prices for the entire metropolis, based on the sampled data, the interpolation technique under spatial analysis tools in ArcGIS software was used. Interpolation predicts values for cells in a raster from limited number of sample data points using a weighted average technique. For the land price interpolation analysis, the Kriging model was adopted. Given that distance partly determined land values in the metropolis, the kriging method was considered appropriate because it considers the spatially correlated distance or directional bias in the data. The Ordinary Exponential Semi-variogram method of Kriging, which assumes that there is no constant mean for the data over an area mean was used (see Childs, 2004). The number of nearest input sample points used to perform the interpolation was set to 12 points, the default number of points. The search limit for the nearest input sample points was set to 3km. The cell size at which the output raster of interpolated land value was created was set at 30.48m × 30.48m resolution to be consistent with the resolution defined for the other layers of the spatial data imported into the model. Output of the interpolated surface, saved as a point-floating grid raster data in ASCII file format was imported into Netlogo using the GIS extension and data loading commands described previously.

7.5.4 Procedure to initializing job market sub-component

The initialization of the job market sub-component of METLOMP-SIM implements two key procedures. As illustrated in figure 7.11, the first procedure takes GIS data on major employment zones as inputs and computes the proximity of locations to each of the employment zones. The output is then loaded into the model and displayed using Netlogo's GIS extension as shown in figure 7.12.

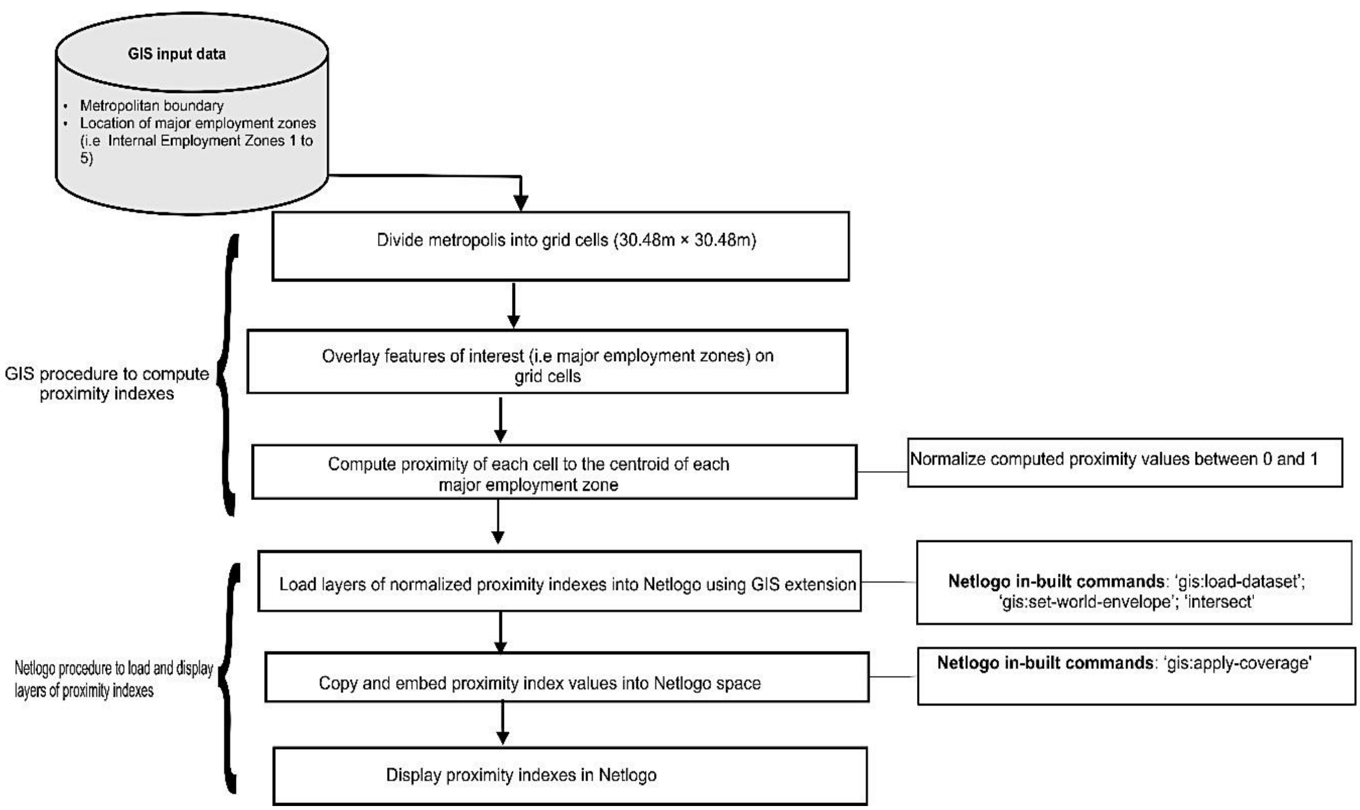


Figure 7.11: Procedure to compute proximity to employment zones and load into Netlogo

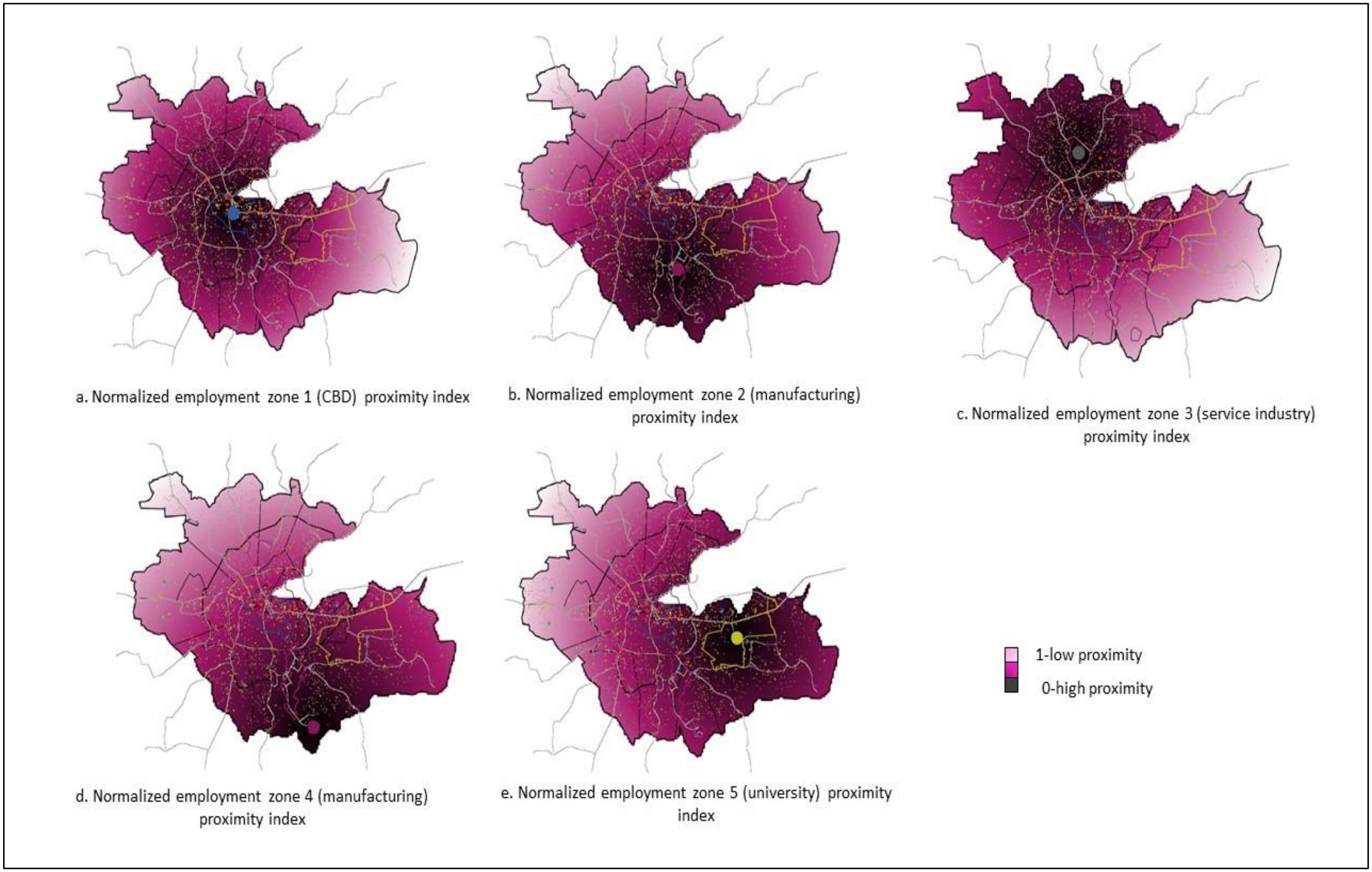


Figure 7.12: Employment zones proximity map layers imported, transformed and visualized in Netlogo using GIS extension

The second procedure of METLOMP-SIM's job market sub-component initialization implements a series of Netlogo commands which generates and distributes an initial number of non-home-based jobs, differentiated by industry and skill requirements among the major employment zones in the model as illustrated in figure 7.13. The characteristics of the job market at model initialization is summarized in table 7.4

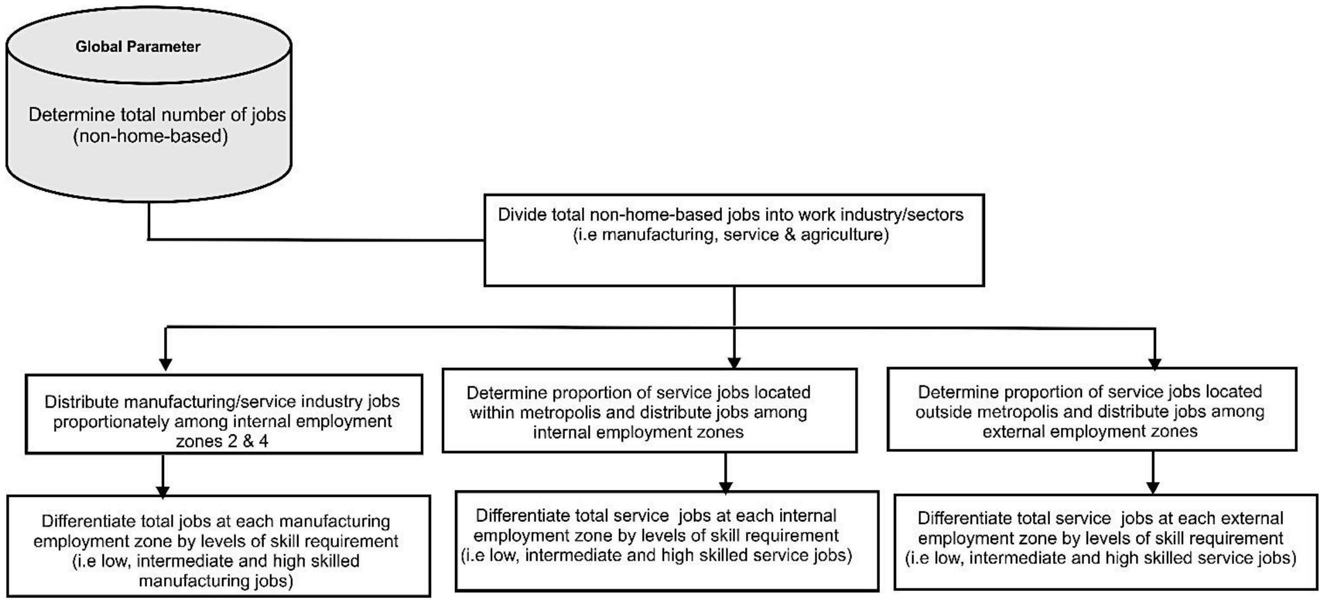


Figure 7.13: Procedure to generate, differentiate and distribute non-home-based jobs among employment zones

Table 7.4: Characteristics of the job market at model initialization

Themes	**Variables**	**Values**
Total jobs	Initial number of jobs (non-home-based)	Specify at initialization
Job industry/ sectors (%)	Service jobs	98
	Manufacturing jobs	1
	Agriculture jobs	1
Skill requirements of jobs (%)	Low-skilled-service jobs	38
	Intermediate-skilled-service jobs	22
	High-skilled-service jobs	40
	Low-skilled-manufacturing jobs	50
	Intermediate-skilled-manufacturing jobs	35
	High-skilled-manufacturing jobs	15
Spatial distribution of service jobs (%)	Service jobs at employment zone-1— Adum/Kejetia CBD	48
	Service jobs at employment zone-2—Asokwa-Ahensan Industrial Enclave	3
	Service jobs at employment zone-3—Magazine Auto-mechanic Enclave	25
	Service jobs at employment zone-4—Sokoban Wood Village	2
	Service jobs at employment zone-5—KNUST, University	7
	Service jobs clustered at markets locations	6
	Service jobs clustered at terminals	3
	Service jobs located outside KMA	6

Source: Based on Field Survey, February 2015

Table 7.4 continued: Characteristics of the job market at model initialization

Themes	Variables	Values
Spatial distribution of manufacturing jobs (%)	Manufacturing jobs at employment zone 2—Asokwa-Ahensan Industrial Enclave	48
	Manufacturing jobs at employment zone 4—Sokoban Wood Village	52
Spatial distribution of service jobs located outside metropolis (%)	External employment zone-1 (Nkawie)	15
	External employment zone-2 (Ejisu)	30
	External employment zone-3 (Mamponteng)	10
	External employment zone-4 (Kodie)	10
	External employment zone-5 (Ofoase-Kokoben)	10
	External employment zone-6 (Asokore Mampong)	20
	External employment zone-7 (Kuntanase)	5

Source: Based on Field Survey, February 2015

7.5.5 Procedure to initialize socio-demographic sub-component

The socio-demographic sub-component of METLOMP-SIM uses the results of the socio-economic characteristics of households derived from the analysis of the survey data to constitute heterogeneous household agents in the model following the steps illustrated in figure 7.14.

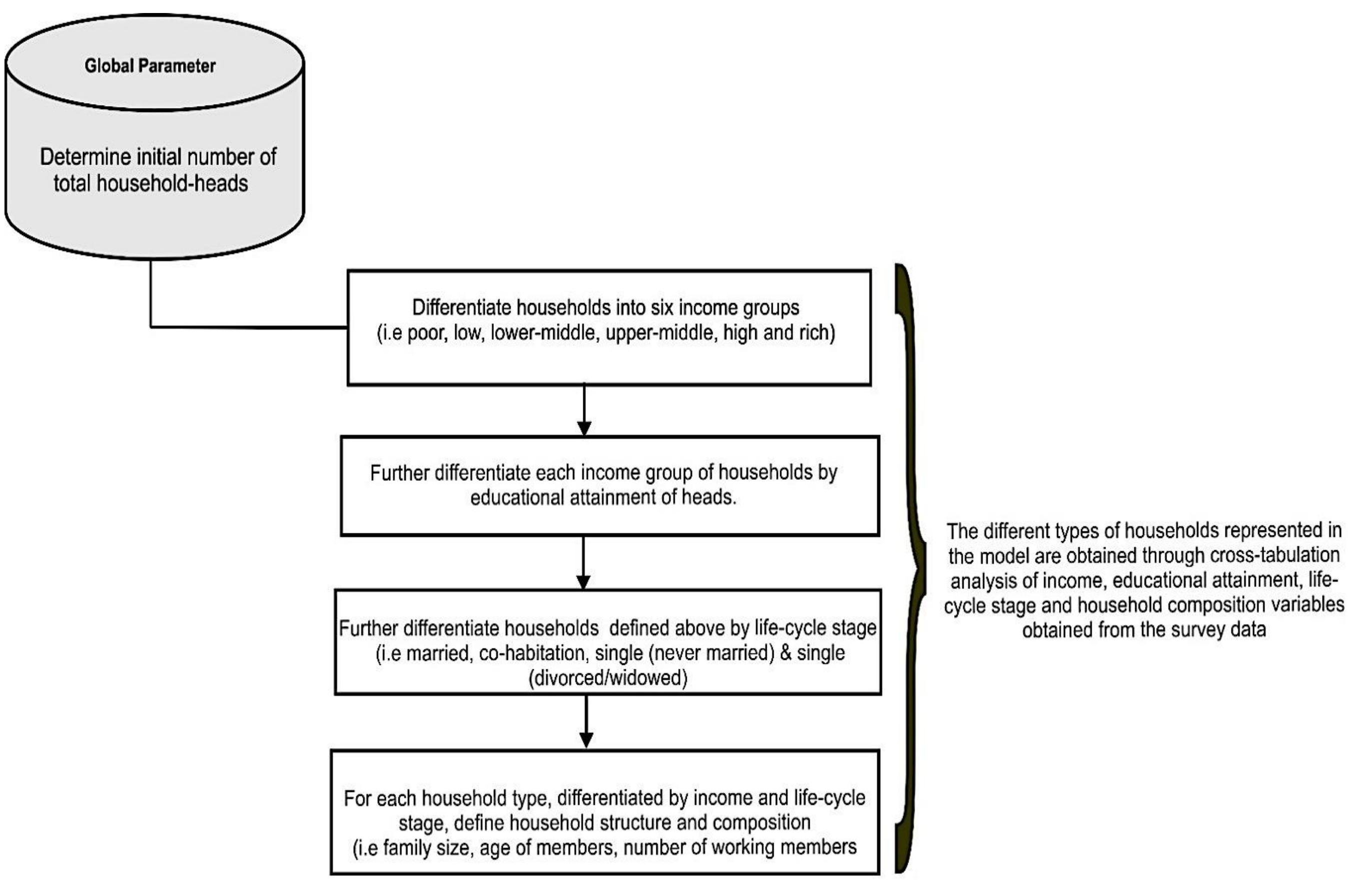


Figure 7.14 procedure to generate and differentiate household agents in METLOMP-SIM

The socio-economic profiles of the households were defined through cross-tabulation of income, educational attainment, life-cycle stage and household composition variables obtained from the survey data. Table 7.5 illustrates the cross-tabulation template used to derive the profiles of differentiated households in the model. The column shows the six income groups while the rows show the other socio-economic variables derived from the survey data.

Table 7.5: Socio-demographic variables cross-tabulation template used to derive different household profiles

Socio-demographic variables	Marital status	Family size	Education/skill of head	Number of working members	Ethnicity	Age distribution
Urban poor						
Low income	*co-habitation*	*3*	*post-secondary, intermediate skill*	*2*	*Akan*	*45yrs-head 30yrs-partner 5 yrs-child*
Lower-middle income						
Upper-middle Income						
High Income						
Rich						

Row three of table 7.5 shows the profile of a typical household represented in the model. This household is defined as a low income, co-habiting household with an intermediate skilled household head, having a family size of three individuals comprising two adult working members aged 45 and 30 years and one child aged 5 years, and belonging to the Akan ethnic group. Adopting this approach, a total of 51 different household profiles was derived from the survey data and reconstituted in the model.

Based on the age profiles of the starting household heads and members of their respective households, new households are endogenously formed during the model execution. During the model simulation, each iteration constitutes one year. A quasi-demographic process is implemented whereby on each iteration, the ages of household members increase by one. At the age of 24 years, which is based on the average marriage age of couples in the observational data, a new household is created in the model. Newly created households inherit the attributes of their original households, and enter the housing and job markets to fulfil their housing and employment location search objectives. This mechanism is responsible for population growth in the model.

The characteristics of the household agents, which reflect the proportion of households differentiated by the various socio-demographic variables at the model initialization are summarized in table 7.6

Table 7.6: Characteristics of household agents at model initialization

Themes	**Variables**	**Values**
Total starting population	Initial households	Specify at initialization
Income groups (%)	Urban poor	1
	Low-income	25
	Lower-middle-income	27
	Upper-middle-income	23
	High income	20
	Rich	4
	18-24	3
Age distribution of household heads (%)	25-34	24
	35-44	30
	45-54	23
	55-64	13
	65+	7
Age distribution of children/dependents of the households (%)	>5	14
	5-9	19
	10-14	22
	15-19	19
	20-24	15
	25-29	8
	30-34	3
	35-39	1
	Low-skilled	47
Skill levels of adult works (%)	Intermediate-skilled	23
	High Skilled	30
	Couple-married	60
Marital status (%)	Single	15
	Co-habitation	10
	Single-divorced/widowed	15
Family size and composition (%)	Family size = 1	15
	Family size = 2	3
	Family size = 3	12
	Family size = 4	51
	Family size = 5	16
	Family size >5	3
	Households with children	82
	Households without children	18
	Working members = 1	29
Working members (%)	Working members = 2	66
	Working members = 3	4
	Working members => 4	1
Ethnic groups (%)	Akan ethnic groups	75
	Northern	17
	Ewe	3
	Ga-Adangbe	1
	Guan	2
	Other ethnic groups	2

Source: Based on Field Survey, February 2015

In summary, the initialization of the socio-demographic sub-component generates the heterogeneous household agents represented in the model. Similarly, the initialization of the spatio-environmental, property market and job market sub-components, generates heterogeneous characteristics of space and discrete spatial goods in the model. The metropolitan landscape consists of cells having a set of variables defining the characteristics of space at a 30.48m × 30.48m resolution—equivalent to the size of a residential land lot in the case study metropolis. Figure 7.15 illustrates the attributes defining space and location when these sub-components of the model are initialized.

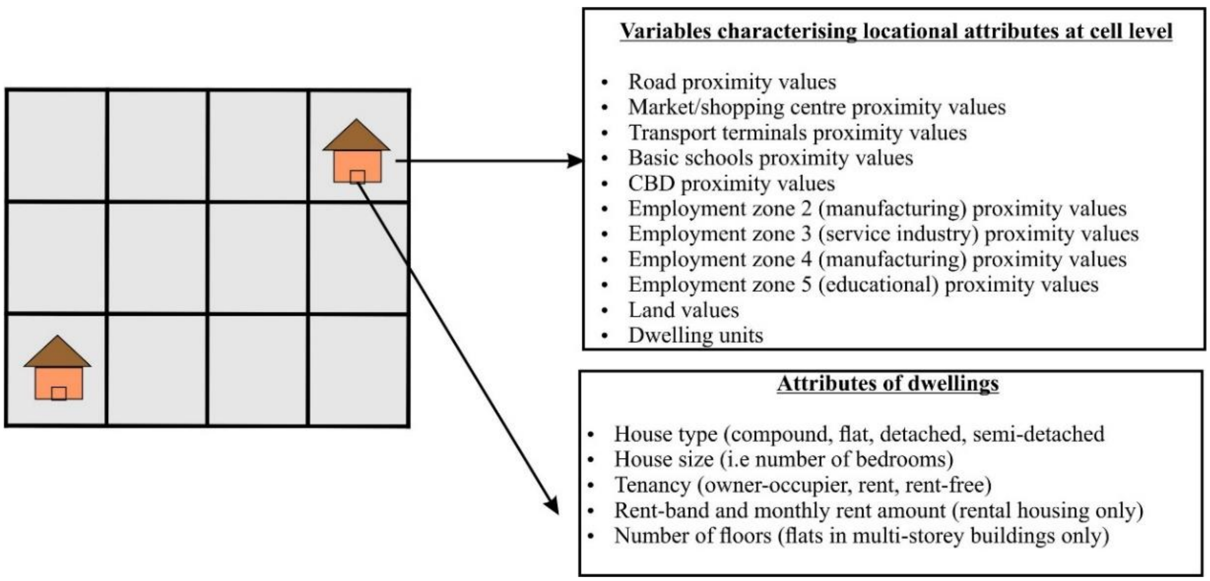


Figure 7.15: Attributes defining heterogeneity of space in METLOMP-SIM

7.6 Execution of METLOMP-SIM: Decision-making tasks, schedules and condition-action-rules

The execution phase of METLOMP-SIM consists of three key procedures programmed to implement sequentially, the decision-making tasks of households and individuals with respect their location and mobility choices as illustrated in figure 7.16. The first encoded procedure, comprising seven sequential tasks, implements the residential location choice behaviour of the household agents. The second procedure at the execution phase contains the sequence of tasks implemented by adult individuals within the households to decide their job location choice. The final procedure implements the travel choice tasks of individuals who have realized their preferred residential and job locations.

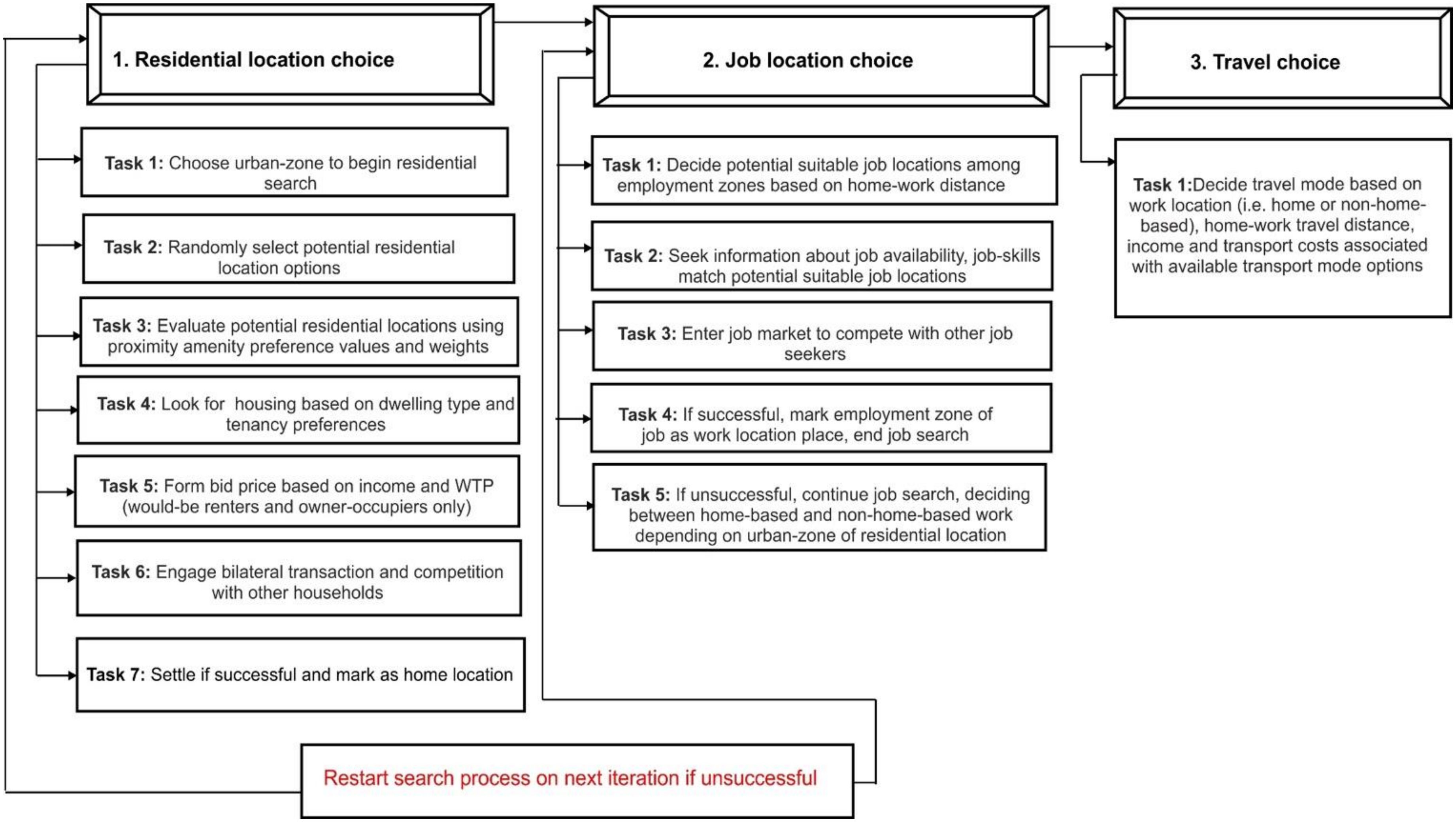


Figure 7.16: Schematic illustration of the model execution procedures and schedule of agents' decision-making tasks

Furthermore, the empirical analysis of residential-job location choice interdependence found that for most of the households in the case study metropolis, residential location decisions preceded their job location decisions. In view of this, a sequential choice process in which agents first choose a place of residence, followed by their employment location choice is implemented in METLOMP-SIM. In the sections that follow, a detailed description of the programmed condition-action-rules and heuristics implemented in METLOMP-SIM under each of the three choice procedures outlined above is provided.

7.6.1 Residential location choice procedure

A schematic illustration of households encoded residential location choice procedure is depicted in figure 7.17. At the start of the model run, household agents are distributed randomly in the urban space to begin their location search process. The residential search process is programmed as sequential tasks explained as follows:

Task 1: Choose urban-zone to begin residential search

From the random initial locations, active household agents begin to search for a place to live by deciding which of the three-broad urban-zones of the metropolis (i.e. historical-core, inner-suburban and outer-suburban zones) to begin their search from—task 1. This is a meso-scale location choice decision taken by household agents with some probabilistic outcomes based on their income groups. The exact values of the probabilities, ranging between 0 and 1 are determined at model calibration. The task of selecting an urban-zone is programmed as a conditional probability rule in the following sequence: The first stage dichotomizes the choice alternatives into two; historical core and suburban locations. Thus, household agents first choose probabilistically between these zones. The second stage of urban-zone choice is conditioned on the outcome of the first. Thus, if historical-core was selected, then the household starts the search from there and ends task one. Otherwise, the suburban zone, divided further into inner-suburban and outer-suburban zones become the choice alternatives at the second stage as shown in figure 7.17.

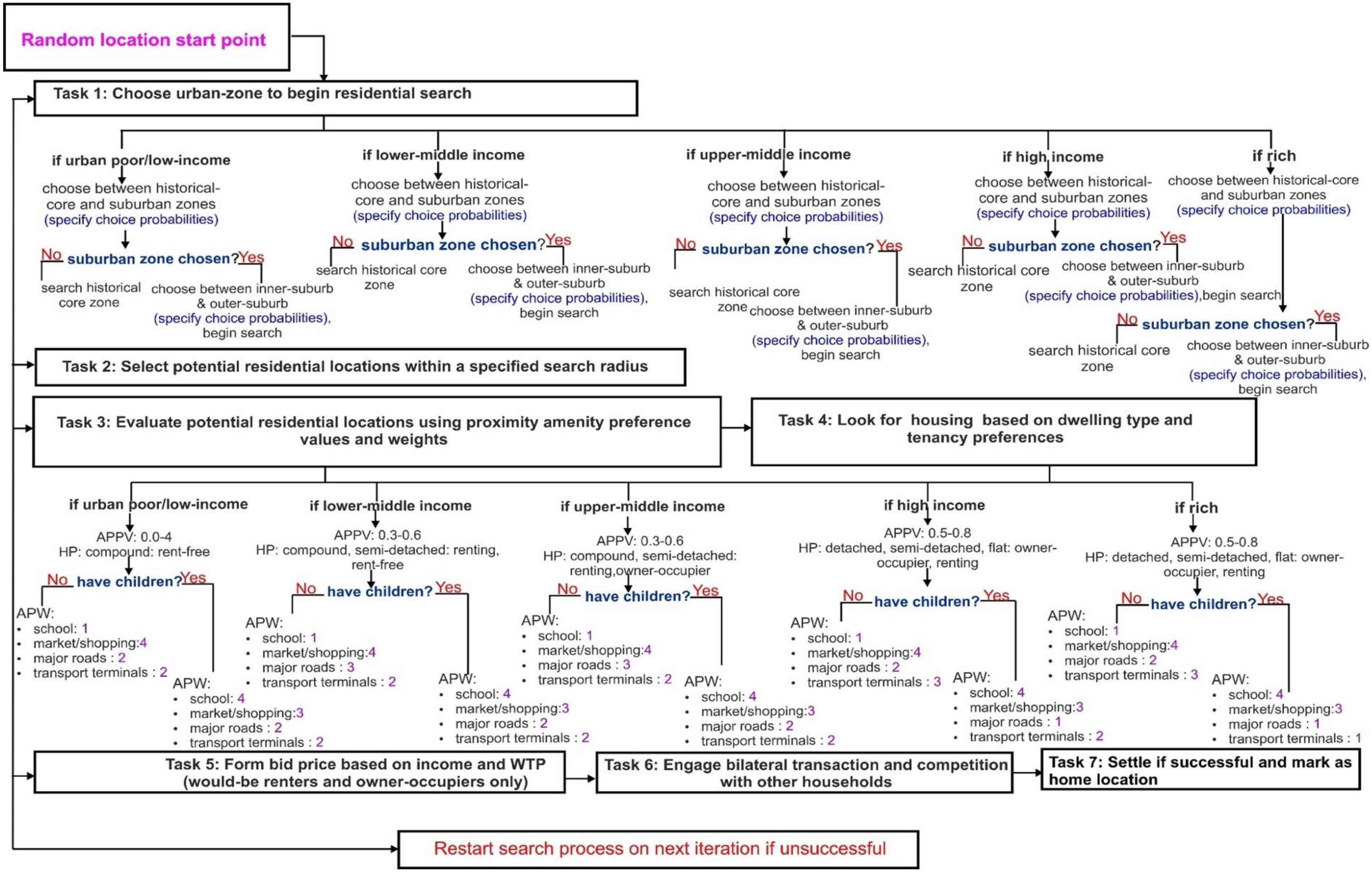


Figure 7:17: Schematic illustration of households' residential location choice tasks schedule and decision-making rules. NB: HP-Housing Preference; APPV-Amenity proximity preference value; APW- Amenity preference weight.

Tasks 2-4: Select potential residential locations within a specified search radius and evaluate based on location proximity and dwelling preferences

From the starting urban-zone, determined in task-1, households implement tasks 2 to 4. As shown in figure 7.17**,** each household type, based on income have housing preference (HP). Low-income households, for example, based on the results of housing preference analysis, have higher preference for compound housing and to live rent-free. Thus, this household samples out of the total dwelling units in the urban-zone being searched within a specified search-radius. The property search radius is a parameter value determined through calibration.

The utility framework used by households to decide their residential locations requires that they possess preference for amenity proximity values (APV), which determines how close or otherwise they want to be in relation to all available amenities. Also, households have amenity preference weights (APW), which priorities the level of importance they attach to each amenity located within the vicinity of their dwelling. APV, being normalized distances between 0 and 1, is categorized and interpreted as follows: 'very high proximity' (0.0-0.2), 'high proximity' (0.2-0.4), 'moderate proximity' (0.4-0.6), 'low proximity' (between 0.6-0.8) and 'very low proximity' (between 0.8-1.0). The household agents rate APW on scale of 1 to 4, where 1 represents low preference and 4 represents very high preference for any given amenity.

The application of the APV and APW by the household agents is demonstrated as follows. As the principal component analysis of households' residential location choice factors in chapter four revealed, proximity to essential amenities constituted one of the major considerations of residential location choice. Moreover, the empirical analysis showed that in general, low income households in the case study metropolis prefer to be closer to these amenities than households of relatively higher income. In view of this, low-income households are assigned APV values of between 0 and 0.4, which is interpreted as high proximity preference as shown in figure 7.17. Similar rules are specified and programmed for households belonging to lower-middle, upper-middle and high income, and rich categories. APVs overlap meaning that there could be similar proximity preferences for amenities across the different household types. This creates one of the conditions for interaction and competition among different agents for locations within the urban area. In terms of weights, each household agents' APW depends, for example, on whether they have children or not. For example, low-income households with children in order of importance, ranks proximity to school their highest priority, followed by proximity to markets/shopping and proximity to major road network and transport terminals (see figure 7.17).

Substituting the APV and APW values into the utility function, the expected utilities of households are compared against utility levels achievable at locations within the households' sample pool as well as the housing preferences are perceived and evaluated. Households do so using satisficing strategy. By the satisficing rule, a household agent does not need to exhaust the complete list of choice alternatives. Instead, households inspect each choice alternative against its preference criteria to select the option that meets their preference.

Moreover, for households to realize their preferred residential locations, adequate number of vacant dwellings should be available in the metropolis. An endogenous mechanism by which the feedback relationship between household population active on the market, the number of vacant properties available and the availability of undeveloped parcels results in the creation of new dwellings is implemented using the heuristic presented in figure 7.18.

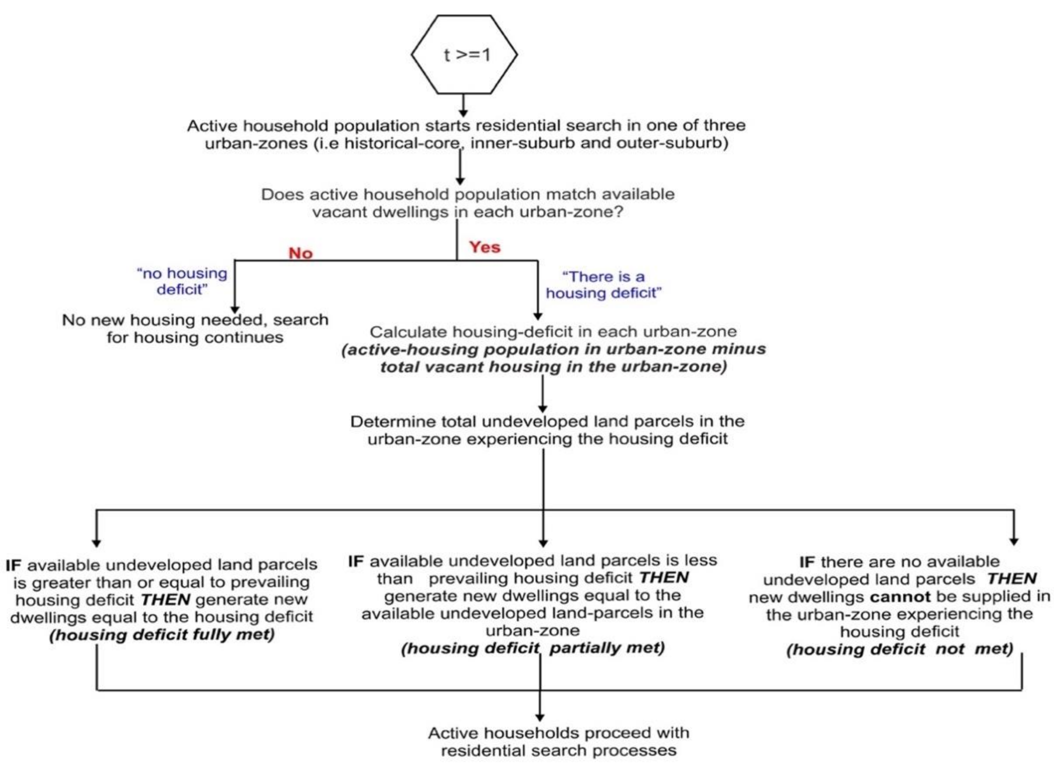


Figure 7.18: Schematic illustration of population-housing supply feedback procedure

The population-housing feedback procedure is implemented at the beginning of each iteration to ensure that housing supply matches closely prevailing demand. The above procedure takes into account vacancy levels in existing housing. Where the number of vacant dwellings is more than or equal to the household population active in the property market, no new houses are provided as this suggests there is either a surplus or a balance between potential housing demand and the existing supply. On the contrary, where there is a deficit—there are no vacancies in existing dwellings and that population actively searching for housing exceeds available units of housing—new housing units are developed on vacant land parcels to accommodate the prevailing demand. In compound housing, since households occupy rooms, vacancy levels are calculated as the number of vacant rooms within existing compound houses. In other housing types where multiple occupation is possible, the number of separate households that could be accommodated in the house is noted at the beginning of the simulation. In subsequent iterations, this number is reduced anytime a new household moves into the house until this number reaches zero, indicating that the house is full.

Having enough housing to accommodate potential demand does not imply that every household active on the market will necessarily be successful in their residential location search. Instead, success depends on the probabilistically determined search process previously outlined, which determines where households begin their search process and which dwellings fall within their sample pool. In the end, the competitive bidding process involving households determining their WTP and submitting bid prices determines residential location outcomes. It is possible that some vacant dwellings will not become the subject of competition in a given iteration. These houses remain vacant and count towards the determination of overall housing need in the subsequent iterations of the model.

Tasks 5-7: Form bid price based on income and WTP, engage in bilateral transactions and competition with other households, and settle if successful

Having found potential suitable residential locations—locations that meet the housing type and tenancy preferences of households as well as their expected utility determined by their preferred amenity proximity values and amenity preference weights, the household agents begin the process of competitive market bidding and bilateral transactions. Households form bid price depending on their willingness to pay (WTP), which is the proportion of their total monthly income they are willing to spend on housing (see figure 7.17, task 5).

As demonstrated through the empirical analysis, the property market of the Kumasi Metropolis, which METLOMP-SIM is intended to replicate, exhibits certain unique characteristics with respect to tenure arrangement and housing development that are different from what might be known in other contexts. The nuances of the property market and how these have been programmed into the agents' behaviour in the model are explained as follows:

Firstly, the empirical analysis of housing tenancy arrangements found the existence of the rent-free housing sector, a non-market housing arrangement linked to the extended family system, offering housing to low-income households. The rent-free housing sector therefore cannot be modelled adopting the conventional market rules of bilateral transactions and competitive bidding arrangements used in representing housing markets. Households in the model are aware of their nuclear family relations but not their extended family relations. In view of these, the rent-free housing market in METLOMP-SIM is modelled using three main considerations derived from the observed relationships between socio-economic characteristics of households and their housing type and tenancy preferences.

The empirical analysis established that most of the households living rent-free do so mainly in traditional compound housing. Moreover, the compound house constitutes one of the prominent and enduring material symbols of extended family relations and support networks in the case study metropolis. Thus, it is reasonable on this basis, to use the compound house and its interaction with the socio-demographic profile of households as the closest proxy to mimic the mechanism by which this non-market sector emerges as a channel of the housing supply in the model. Using these empirically grounded principles, households who opt for living rent-free do not go through the process of forming bid price, submitting bids and engaging in competitive bidding. Instead, once their socio-economic circumstances and preferences lead them into choosing this form of tenure, the only condition that must be met is that there is vacancy for a rent-free tenant within one of the properties they have evaluated. If this condition is satisfied, the household settles, ending the residential search process.

The rental housing market on the other hand, lends itself to conventional property market modelling. Households who prefer rented housing therefore decide how much they are willing to spend on renting the accommodation (i.e. WTP)—see figure 7.17, task 5. WTP for each household type, is derived and calibrated empirically from the survey data analysis. Within each broad income group, monthly earnings are further divided into two, the lower-half and upper-half of the earnings. For each half, the empirically derived rent-to-income affordability threshold represents the households' WTP.

Following the property market principle that households intend to benefit from the surplus of trade, they do not submit the full WTP at the first stage of competition. Rather, they submit an initial bid price which is less than five percent of their WTP. The submitted bid prices are compared with the ask price of the targeted properties. The outcome of the market competition is determined by the highest bidding household. If there are more than one household submitting the highest bids, then one of them is chosen randomly as the winner. In the subsequent iterations, households who were not successful during the previous round of search re-enters the market, this time, bidding the full WTP with the objective to increase their chances of renting a house at their preferred location. Similarly, dwellings which after being evaluated by the households more than a given number of times remains on the market unlet, reduces its ask rent amount in response to the demand side of the market.

The bilateral transactions and competition between households in the rental market accounts for the formation and evolution of prices endogenously in the model. A schematic illustration of the price formation and evolution heuristics in the rental market is presented in figure 7.19.

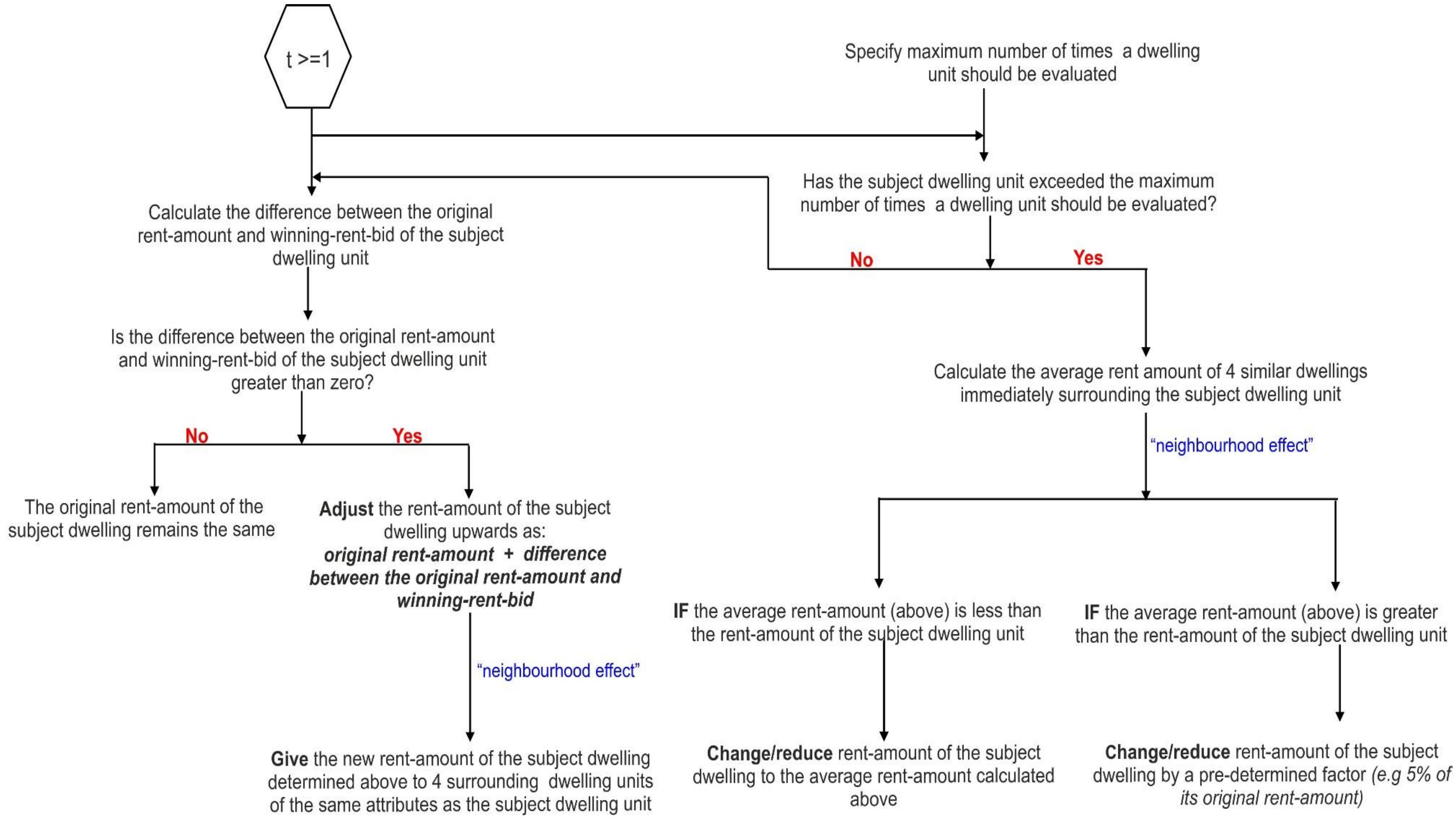


Figure 7.19: Schematic illustration of rental market price adjustment and formation heuristic

The last market tenure choice simulated in METLOMP-SIM is the owner-occupier tenancy arrangement. Similar to the rent-free housing sector, the owner-occupier housing market of the case study metropolis which the model represents, deviates from the conventional market principles, requiring a somewhat different decision-making heuristic. Within the context of the KMA, as indicated by the survey data analysis, most owner-occupiers acquire land and develop their own housing incrementally over several years. A detailed representation of the incremental housing development process, however, is not required to realize the objectives of the current model, and is therefore not explicitly modelled. Instead, the emergent location outcomes of owner-occupiers are modelled using a set of heuristics that mimic the market behaviour of households in the owner-occupier sector. This is accomplished in METLOMP-SIM by using the stock of complete owner-occupier dwelling units generated on each iteration of the model to represent the potential locations where would-be owner-occupier households could find suitable housing. The parcel of land on which the house is located instead, has the reservation price and forms the object of choice and market competition. Using this approach, the focus of households' market behaviour is shifted to the parcel of land occupied by the dwelling unit while the dwelling unit itself is used to represent the attributes of the home that the households would have eventually built.

In addition, as evidenced by the survey results, would-be-owner-occupiers tend to accumulate a lump sum often higher than their monthly incomes to purchase land. In view of this, households entering the land market are assumed to have a lump sum amount equivalent to a minimum of 5 years and a maximum of 10 years of their total monthly earnings. This lump sum amount constitutes their WTP, which they submit to purchase the land at their preferred residential locations.

Similar to the rental housing market, outcomes of competition for land among would-be-owner occupiers is determined by the highest bidder. Households that are not successful re-enter the market on subsequent iterations to begin their residential search process. The bilateral transactions and competition among households in the land market also accounts for the formation and evolution of land prices endogenously in the model. A schematic illustration of the land price formation and evolution heuristics is presented in figure 7.20.

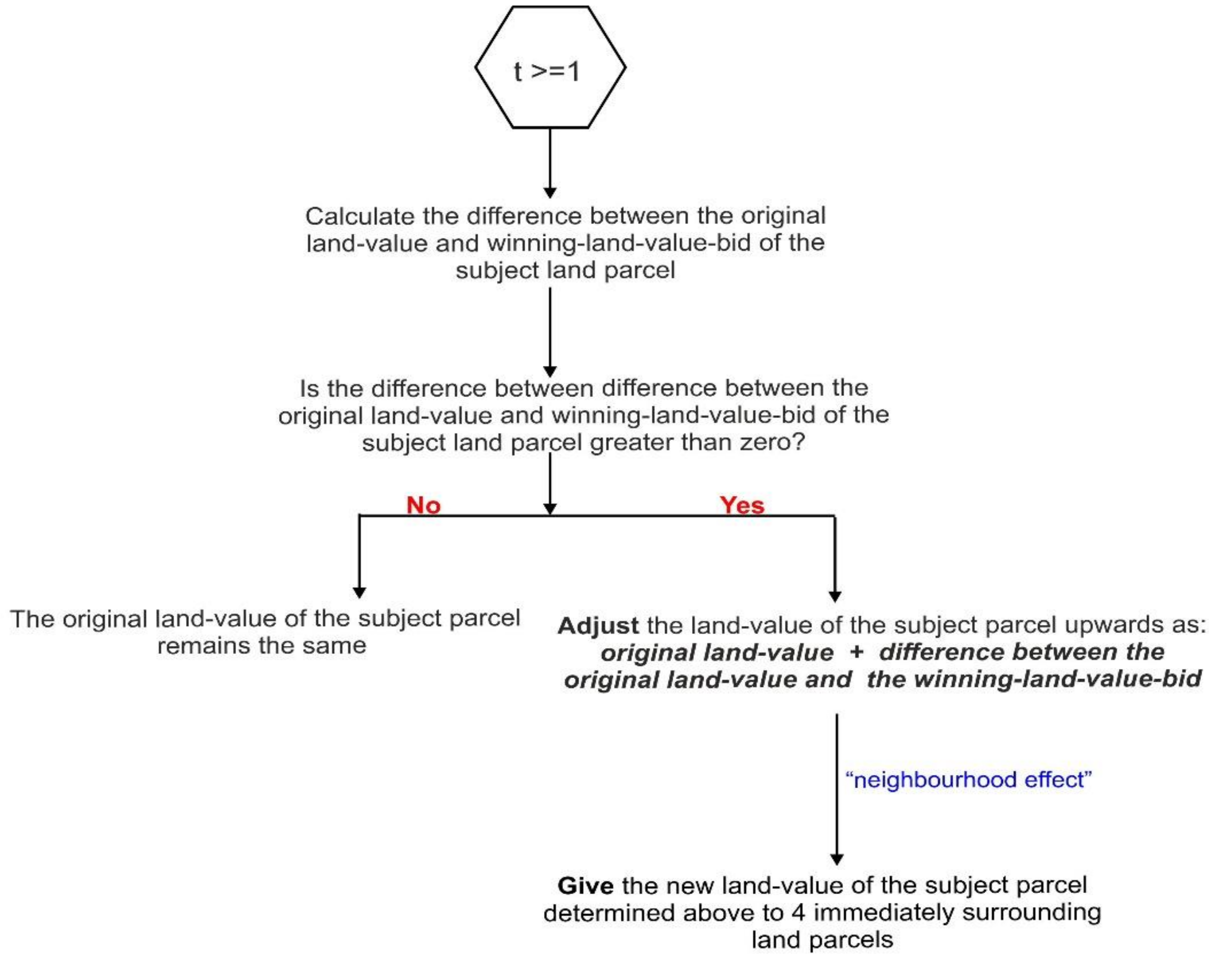


Figure 7.20: Schematic illustration of land market price formation and evolution heuristics

In summary, the bilateral transactions represented in the model allows for intra-group and inter-group competition. For example, although households within the same income-group may have similar housing preferences, their WTP and bid price may not necessarily be the same. This implies that within the same income group, different households may have slightly different market abilities/powers which constrain their preferences. Moreover, it is possible for a given location to meet the minimum amenity-proximity preference criteria of households but the intrinsic attributes of the dwelling there would not meet the households' preferences. Since the latter is very important condition that should be met, a household faced with this scenario would have to evaluate another property in the subsequent iteration of the model.

7.6.2. Job location choice procedure

In METLOMP-SIM, job location choice follows sequentially the residential location choice of households. Thus, from the residential locations realized by the households, individual adult members within the household make job location decisions. A schematic illustration of the job location choice procedure is presented in figure 7.21. The five main tasks that constitute the job search process are discussed in the sections that follow.

Task 1-2: Decide potential suitable job locations among major internal employment zones and seek information about job availability and job-skills match

Individual job seekers start their job location choice process from the attained home locations of their respective households. The first task of the programmed procedure involves job seekers deciding a potential suitable job location among the five major employment zones located within the metropolis. Following findings of the empirical analysis, job seekers choose one of the internal major employment zones with the minimum distance from their home location to look for a job. Next, the prospective job seeker, moves to the selected potential employment zone to obtain information about job availability and to ascertain if the available job matches their skills.

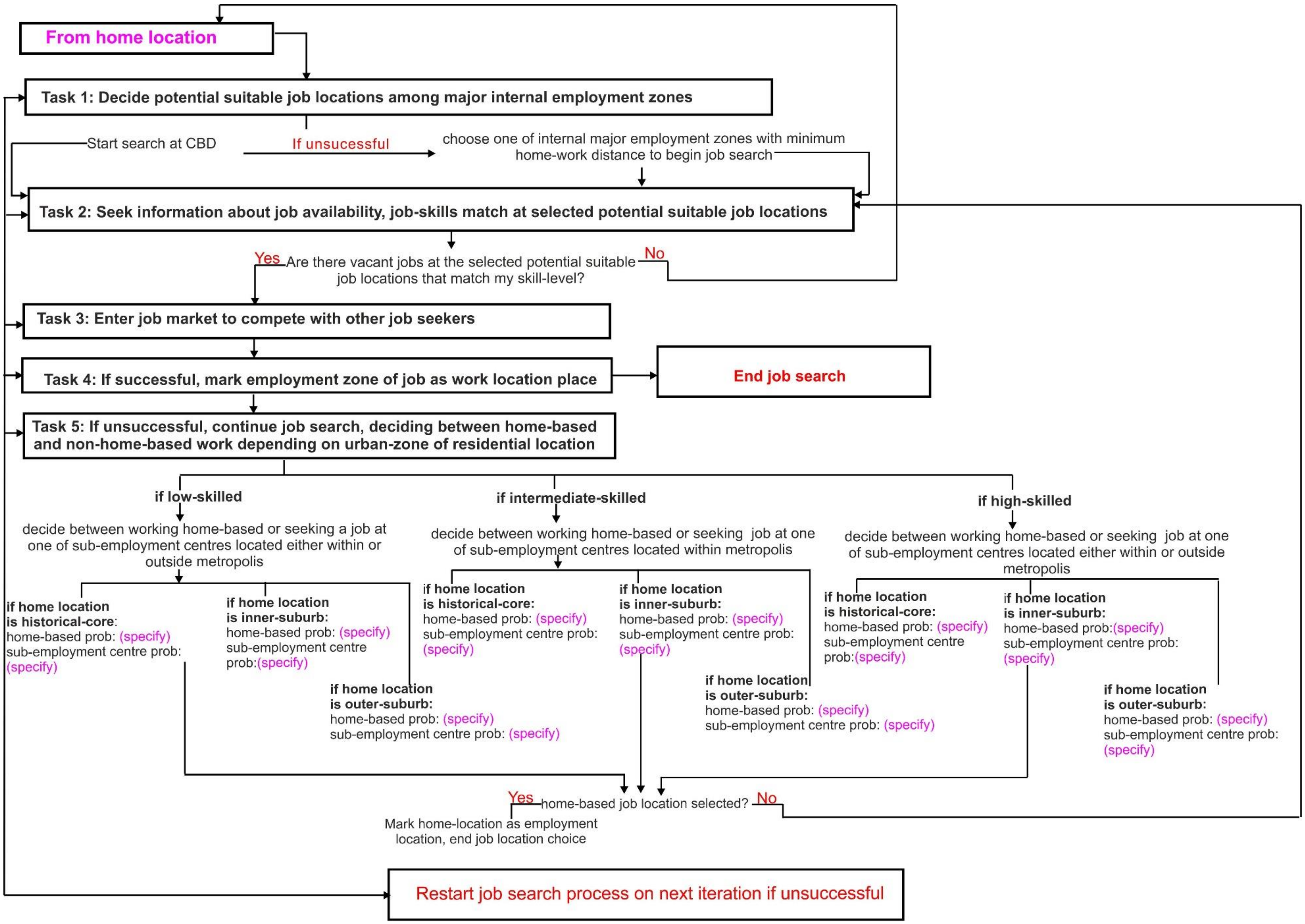


Figure 7.21: Schematic illustration of job location choice procedure

In the model, the job market responds to the number of job seekers at each of the major employment zones in the metropolis. This feedback mechanism, illustrated in figure 7.22 ensures that new jobs of the required skills are generated endogenously over time to match the growing number of job seekers.

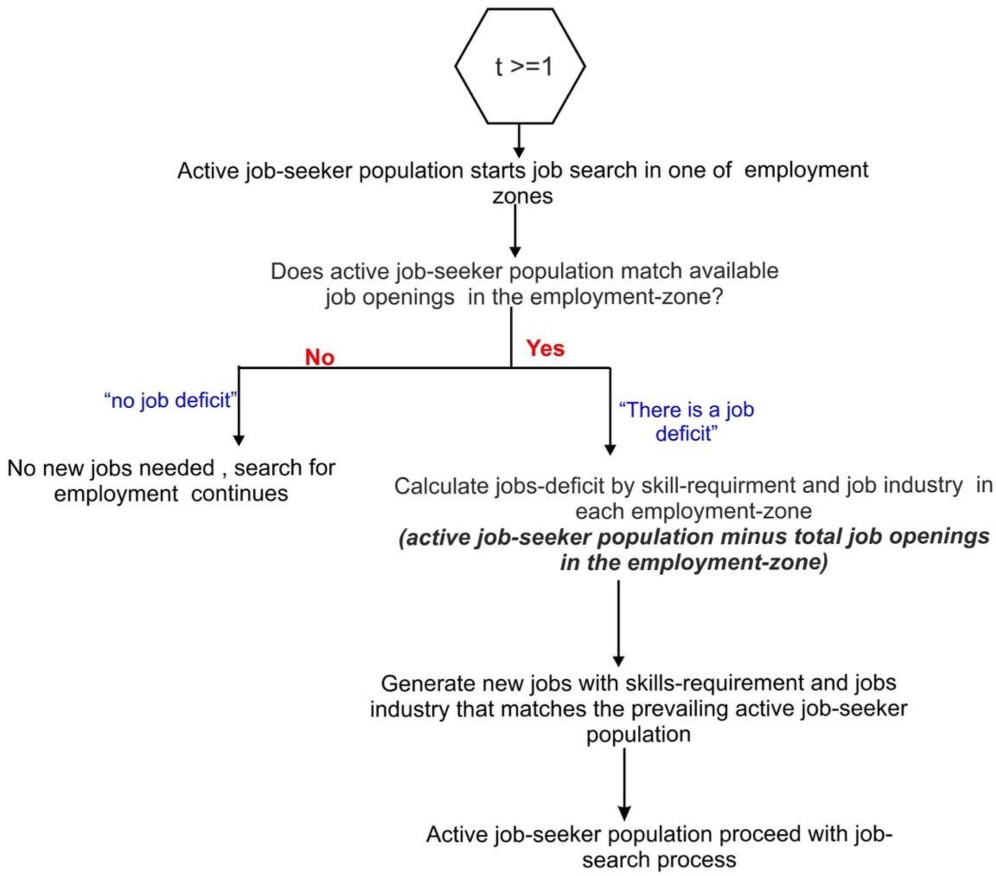


Figure 7.22: Schematic illustration of population-jobs supply feedback heuristics

Task 3-5: Enter job market to compete with other job seekers

Once a vacant job which matches the skill-level of the prospective job seeker is identified at the selected potential suitable job location as explained above, they make the decision to compete with other job seekers for the available job openings. The outcome of the job search competition is determined by the number of job openings at the potential job location, the number of prospective job seekers competing for that job opening and the fit between the candidates’ skills and the skill requirement of the available job(s). In instances where more than one job seeker competes for a single job opening that matches their skill levels, one of them is chosen randomly and offered the job. The successful candidate marks the employment zone where the job offer was made as their place of work or employment location and ends the job location choice process, establishing their home-work location combinations.

Prospective job seekers, also choose between home-based job location and non-home-based job location. This occurs in the same time iteration of the simulation. The decision to either opt for home-based employment or non-home-based employment occurs endogenously. Following the findings of the empirical study of job location choice in the Kumasi metropolis, the choice between home-based and non-home-based job locations is determined probabilistically depending on the prospective job seekers' skill level and urban-zone of home location. The appropriate values of these choice probabilities for each of the prospective workers in the three urban zones are determined at the model calibration stage. Individuals who are unsuccessful at securing a job after going through all the stages of the job search process illustrated in figure 7.21 re-enters the job market on subsequent iterations of the simulation to begin the job search process.

7.6.3. Travel choice procedure

The final choice procedure programmed in METLOMP-SIM involves individuals who have their residential and job locations realized making their home-work travel choice decisions. A schematic illustration of the travel choice procedure, which implements travel mode choice rule is summarized in figure 7.23.

Individual workers choose between two modes of transport namely; active (i.e. walking) and motorized modes of transport based on empirically determined probabilities assigned at calibration. If the individual worker's job location is home-based, then following the findings of the empirical analysis of work travel mode choice, the preferred travel mode would be walking. On the contrary, workers opt for motorized transport, whether private or public if their work locations are non-home-based. The choice between private car and public transport as work travel mode is determined probabilistically based on the income group of the commuter. Similarly, public transport users choose probabilistically between two alternatives- Mini-bus/Trotro and Taxi as depicted in figure 7.23

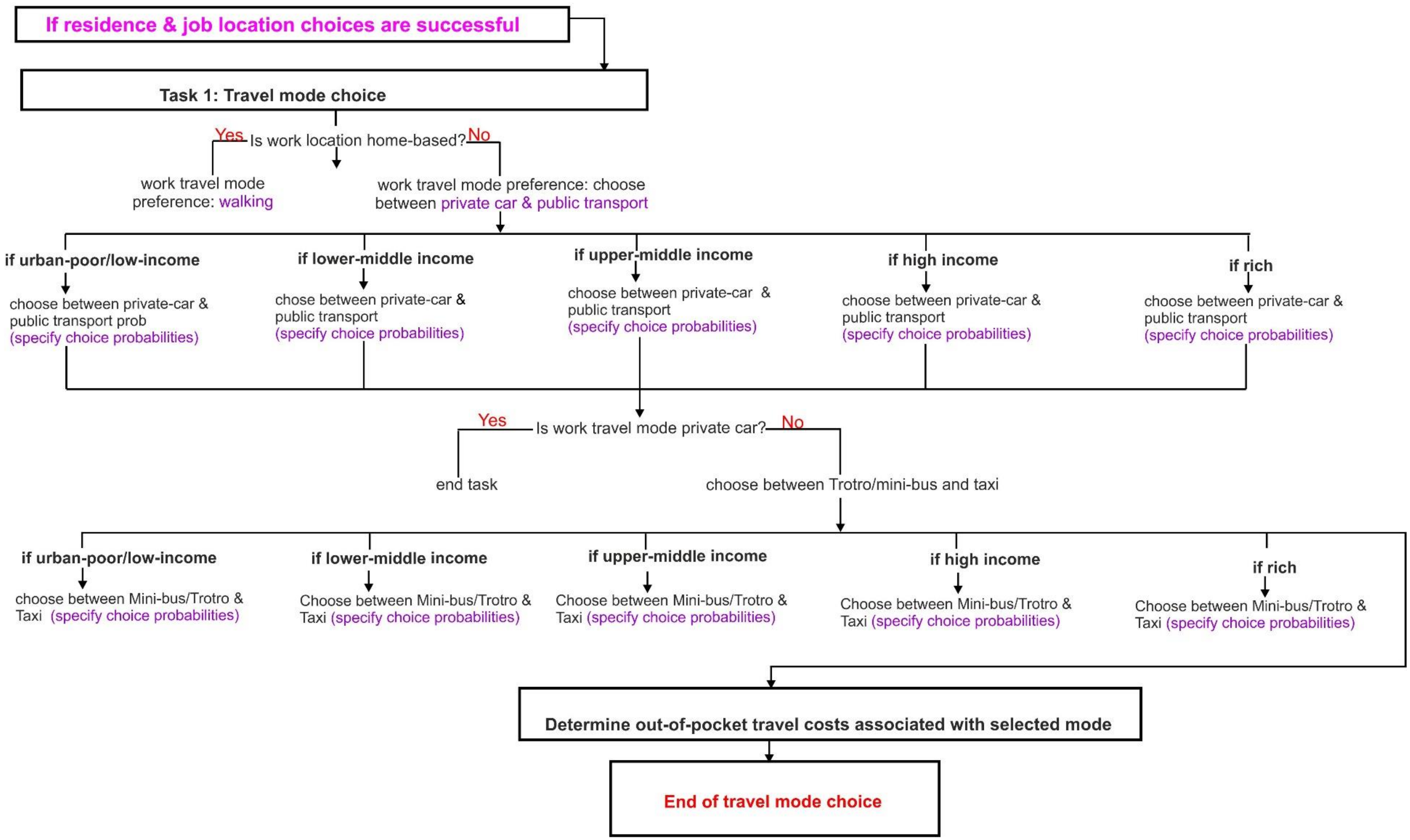


Figure 7.23: Schematic illustration of the choice procedure

7.7 Chapter Summary

Using the conceptual model presented in chapter six and interpreting results of the data analysis presented in chapters four and five, this chapter has provided a detailed discussion of the encoded processes and procedures implemented in the integrated agent-based and geospatial model of urban location choice and mobility patterns. By situating the conceptual model within the unique context of the Kumasi metropolis, the case study area for this research, this chapter has demonstrated in detail, how the empirical insights have informed the formulation of agents' rules of behaviour and decision-making heuristics, and how these have been programmed in METLOMP-SIM.

As indicated at the beginning of the chapter, the model which was to be implemented recognised the urban location and spatial interaction nexus as a function of the interaction between the location choice behaviour of heterogeneous households and individuals in the urban property and job markets, and existing urban structural conditions. With this in mind, this chapter proceeded to offer a detailed discussion of the encoded procedures and commands responsible for generating the initial conditions in the model. This was done for each of the four sub-components, comprising the spatio-environmental, property market and socio-demographic sub-components.

Moreover, the processes involved in three main sequential choice tasks executed by agents in the model namely; residential location choice, job location choice and work travel mode choice were outlined. The overall schedule of the tasks as well as the programmed heuristics and condition-action-rules underpinning performance of the tasks by the agents in the model were also outlined. The encoded feedback mechanisms between population growth and how this generates demand on the one hand, and housing and jobs growth, representing supply on the other hand were also discussed. Finally, the nature of competition between agents in the property and job markets as well as the adaptive learning mechanisms used by the purposive agents to realize their objectives.

The programmed procedures and commands presented in this chapter is the precursor to the next step of the model development process presented in chapter eight, which focuses on the model calibration with data from the case study context and the analysis of results of simulations.

CHAPTER EIGHT: MODEL CALIBRATION EXPERIMENTS AND SIMULATION RESULTS

8.1 Introduction

In chapter seven, a detailed discussion of the encoded procedures, condition-action-rules and decision-making heuristics implemented in METLOMP-SIM—an integrated geospatial and agent-based model to simulate the co-emergence of urban location choice and mobility pattern, was presented. Building on the model implementation processes presented in chapter seven, this chapter focuses on the calibration of the programmed model for the Kumasi Metropolis, the cases study area of this research as well as the results of the simulations generated for the case study metropolis.

The model calibration process involves a series of carefully designed experiments by which the minimum model run/repetitions is established and the parameter space of the model explored to arrive at combinations of parameter settings that produce simulated outputs of interest consistent with the observational data. The best-fit parameter settings of the model are then used to generate a final simulation from which the model's outputs of interests are interpreted.

This chapter is organized into three main sections. In the first section, the model parameter estimation and calibration experiments conducted to arrive at the set of best-fit parameter settings are discussed. This is followed with a discussion of the simulation results of the calibrated model, focusing on the emergent residential location patterns of the simulated households and the formation and evolution of price in the land and housing markets, as a result of the bilateral transactions and competition among the household agents. In addition, the emergent employment location patterns of individual working members of the households, and the emergent home-work mobility patterns in terms of simulated home-work distance separation, work trip production and attraction patterns and work travel mode choice are discussed. In the concluding section, a summary of the model simulation results is provided.

8.2 Model parameter estimation and calibration experiments

The model was calibrated by systematically adjusting the parameter values to find the set of values that enables it to reproduce the characteristics of metropolitan context being modelled. Fine-tuning the model parameters in a systematic and transparent manner involved undertaking a five-stage simulation experiment. This involved identifying calibration parameters and initial parameter values; determining the minimum number of simulation runs/repetitions; deciding the type of calibration to conduct and defining the calibration criteria; exploring the model parameter space to determine optimal parameter values; and running the final model with the best parameter settings (see e.g. Railsback and Grimm, 2011; Thiele et al., 2014; Lee et al., 2015).

8.2.1 Setting initial parameter values and determining minimum model simulation repetitions

In table 8.1, all the parameters of the three main choice processes implemented in METLOMP-SIM are identified. Initial values of the parameters were derived from the fine-scale empirical data obtained through the cross-sectional survey and analysed in chapters four and five of this thesis. These initial parameter values are referred to as 'empirical-values'.

Having identified the model parameters and estimated their initial values from the observational data, the next stage of the calibration experiment focused on determining the minimum number of model simulation run/repetitions. This step is critical because of the stochastic effects and underlying complex dynamics in the model which affect the certainty of the model outputs. Moreover, given that the model is computationally expensive, determining the minimum number of repetitions helps avoid either producing too many sample runs or producing too few sample runs to the extent that the simulation outputs harbour great uncertainty.

Table: 8.1: Model parameters and initial parameter values derived from the observational data.

Choice Process	**Parameter Name**	**Parameter value settings**
		Empirical-values
Residential choice: Urban zone choice probabilities ***(Range: 0-1)***	Core/suburban choice probability: urban-poor	0.38
	Inner-suburban /outer- suburban choice probability: urban-poor	0.65
	Core/suburban choice probability: low-income	0.38
	Inner-suburban /outer- suburban choice probability: low-income	0.65
	Core/suburban choice probability: lower-middle-income	0.29
	Inner-suburban /outer- suburban choice probability: lower-middle-income	0.55
	Core/suburban choice probability: upper-middle-income	0.22
	Inner-suburban /outer- suburban choice probability: upper-middle-income	0.51
	Core/suburban choice probability: high-income	0.17
	Inner-suburban /outer- suburban choice probability: high-income	0.63
	Core/suburban choice probability: rich	0.17
	Inner-suburban /outer- suburban choice probability: rich	0.63
Residential choice: WTP for housing in rental market (% of monthly income) ***(Range: 0-100)***	WTP Low-income (lower-half)	11%
	WTP Low-income (upper-half)	15%
	WTP Lower-middle income (lower-half)	10%
	WTP Lower-middle income (upper-half)	12%
	WTP Upper-middle income (lower half)	32%
	WTP Upper-middle income (upper-half)	55%
	WTP High income (lower-half)	20%
	WTP High income (upper-half)	31%
	WTP Rich (lower-half)	50%
	WTP Rich (upper-half)	60%
Job choice: Home-based employment choice probability ***(Range: 0-1)***	Home-based employment choice probability: low-skilled at historical-core zone	0.31
	Home-base employment choice probability: low-skilled at inner-suburban zone	0.23
	Home-base employment choice probability: low-skilled at outer-suburban zone	0.30
	Home-base employment choice probability: intermediate-skilled at historical-core zone	0.38
	Home-base employment choice probability: intermediate-skilled at inner-suburban zone	0.25
	Home-base employment choice probability: intermediate-skilled at outer-suburban zone	0.31

Table 8.1. continued: Model parameters and initial parameter values derived from the observational data.

Choice Process	Parameter Name	Parameter value settings
		Empirical-values
Job choice:	Home-base employment choice probability: high-skilled at historical-core zone	0.32
Home-based employment choice probability	Home-base employment choice probability: high-skilled at inner-suburban zone	0.23
(Range: 0-1)	Home-base employment choice probability: high-skilled at outer-suburban zone	0.26
Travel choice:	Private car ownership probability: urban poor	0.01
Motorized travel mode choice	Private car ownership probability: low-income	0.01
(private car vs public transport)	Private car ownership probability: lower-middle-income	0.08
(Range: 0-1)	Private car ownership probability: upper-middle income	0.25
	Private car ownership probability: high income	0.57
	Private car ownership probability: rich	0.78
Travel choice:	Mini-bus/Trotro choice probability: urban poor	0.91
Public transport mode choice	Mini-bus/Trotro choice probability: low-income	0.88
(mini-bus/Trotro vs taxi)	Mini-bus/Trotro choice probability: lower-middle-income	0.91
(Range: 0-1)	Mini-bus/Trotro choice probability: upper-middle income	0.88
	Mini-bus/Trotro choice probability: high income	0.79
	Mini-bus/Trotro choice probability: rich	0.64

To determine the minimum sample runs, the initial empirical values of the model parameters outlined in table 8.1 were used to run the model several times. Defined outputs of interests were then examined for their variability using the Co-efficient of Variation (CV) metric. CV is the ratio of the standard deviation to the mean of each of the model output of interest. This is depicted in equation 8.1, where δ is standard deviation of model output of interest and μ is the mean of the sample outputs of interest.

$$\mathrm{CV} = \frac{\delta}{\mu} \qquad 8.1$$

CV provides a meaningful way to compare the difference between model outputs over several sample runs/repetitions. The aim is to determine the number of repetitions where the CV values of the simulation outputs stabilizes and remain the same; the point where the CV converges is considered a minimum sample size or minimum number of ABM runs (Lorscheid et al., 2012; Thiele et al., 2014).

Using the empirical values of the model parameters, the model was run multiple times beginning with five repetitions and increasing the number sequentially in the increment of five. Each simulation initialized with a total household population of 10,000. After each run, CV metrics for all the model output of interest were computed. Table 8.2 provides the CV measures of critical model outputs computed between the repetition samples of five through to 50. The analysis shows that overall, the CV measures of the simulation output converges around 15 repetitions on the low side and 45 repetitions on the high side. Thus, a minimum run of 45 repetitions per simulation is required to consider the stochastic effects in the model and to reduce uncertainty resulting from variations in output of interest.

It is important to indicate that, the purpose of the calibration experiment at this stage was to determine the minimum number of model run but not to ascertain whether the outputs fit the observational data or not. Comparing the model results to some predetermined calibration criteria is discussed in the sections that follow.

Table 8.2: Co-efficient of variation estimates for model outputs of interest across simulation repetitions

Simulation output of interest	Number of runs/repetitions									
	5	10	15	20	25	30	35	40	45	50
Total simulated households										
Mean	19862	19875	19843	19853	19849	19848	19849	19852	19854	19853
Co-efficient of variation	0.001	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003
Renters										
Mean	9655	9651	9634	9658	9665	9668	9668	9673	9681	9685
Co-efficient of variation	0.014	0.011	0.011	0.012	0.013	0.013	0.012	0.012	0.012	0.012
Owner-occupiers										
Mean	708	715	718	715	720	722	722.	721	720	719
Co-efficient of variation	0.026	0.025	0.030	0.027	0.030	0.029	0.027	0.027	0.026	0.026
Rent-free tenants										
Mean	3846	3858	3867	3849	3841	3834	3826	3819	3819	3813
Co-efficient of variation	0.026	0.021	0.023	0.026	0.029	0.031	0.031	0.031	0.030	0.029
Historical-core zone of residence										
Mean	6638	6619	6601	6599	6587	6579	6571	6572	6580	6580
Co-efficient of variation	0.011	0.015	0.019	0.018	0.018	0.021	0.020	0.019	0.018	0.018
Inner-suburban zone of residence										
Mean	3995	4019	4036	4037	4047	4058	4055	4055	4052	4051
Co-efficient of variation	0.036	0.036	0.037	0.037	0.035	0.038	0.037	0.035	0.034	0.032
Outer-suburban zone of residence										
Mean	3551	3562	3559	3563	3570	3563	3567	3562	3565	3563
Co-efficient of variation	0.015	0.011	0.014	0.014	0.019	0.024	0.027	0.026	0.025	0.025
Average distance of home to school (households with children only)										
Mean	556.28	550.61	552.43	551.91	553.96	553.64	553.70	553.74	553.81	553.93
Co-efficient of variation	0.007	0.015	0.014	0.015	0.023	0.022	0.021	0.020	0.020	0.020
Average distance of home to shopping										
Mean	1498.63	1490.16	1490.54	1490.21	1494.23	1493.75	1493.79	1493.64	1493.67	1493.54
Co-efficient of variation	0.006	0.010	0.009	0.011	0.012	0.011	0.012	0.011	0.011	0.011

Table 8.2: continued: Co-efficient of variation estimates for model outputs of interest across simulation repetitions

Simulation output of interest	Number of runs/repetitions									
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**
Average distance of home to transport terminal										
Mean	1643.46	1641.40	1638.91	1641.87	1650.08	1649.09	1650.76	1648.66	1648.44	1648.08
Co-efficient of variation	0.009	0.010	0.011	0.011	0.017	0.017	0.018	0.018	0.018	0.018
Average distance of home to major roads										
Mean	404.22	403.35	404.31	404.33	404.34	404.63	404.84	404.87	404.36	404.62
Co-efficient of variation	0.010	0.009	0.012	0.014	0.015	0.015	0.014	0.014	0.014	0.014
Total non-home-based employment locations										
Mean	8443	8453	8435	8434	8434	8448	8443	8444	8457	8465
Co-efficient of variation	0.013	0.014	0.015	0.014	0.014	0.014	0.014	0.013	0.013	0.013
Total home-based employment locations										
Mean	10393	10446	10468	10481	10499	10481	10483	10468	10460	10442
Co-efficient of variation	0.010	0.012	0.012	0.014	0.015	0.014	0.014	0.014	0.015	0.016
Average home-work distance										
Mean	4641.07	4622.36	4616.59	4617.75	4621.90	4617.84	4620.69	4621.37	4620.85	4621.02
Co-efficient of variation	0.016	0.013	0.013	0.013	0.013	0.012	0.012	0.012	0.012	0.012
Motorized transport use										
Mean	8455	8466	8447	8447	8447	8461	8456	8457	8470	8477
Co-efficient of variation	0.010	0.012	0.012	0.014	0.015	0.014	0.014	0.014	0.015	0.016
Non-motorized transport users (walking)										
Mean	10411	10463	10485	10499	10517	10499	10500	10485	10477	10458
Co-efficient of variation	0.013	0.013	0.015	0.014	0.014	0.014	0.013	0.013	0.013	0.013
Private-car ownership and use										
Mean	1279	1290	1291	1285	1283	1281	1279	1284	1289	1287
Co-efficient of variation	0.029	0.037	0.031	0.033	0.034	0.034	0.033	0.034	0.034	0.034
Public transport use										
Mean	7176	7175	7156	7162	7164	7179	7176	7172	7181	7190
Co-efficient of variation	0.016	0.014	0.015	0.016	0.015	0.015	0.014	0.014	0.014	0.014

8.2.2 Calibration method and parameter-sweeping experiments

In general, ABMs can be calibrated using two main methods namely; categorical and best-fit calibration. Whereas the former requires searching for parameter values that produce outputs of interest within a range of plausible values based on the observational data, the latter requires setting some specific criteria and identifying the set of parameters that produce results that match exactly this criterion (Railsback and Grimm, 2012; Lee et al., 2015).

METLOMP-SIM was calibrated using the categorical method. This method was considered appropriate over best-fit calibration mainly because the observational data being used for the calibration was obtained through a cross-sectional survey of a random sample of households and individuals. Sampling errors and uncertainties as well as issues of variability, representativeness and generalizability, inherent in survey data imply that deriving single measures as calibration criteria from them would be inappropriate. In addition, the complexity of ABMs and the underlying stochasticity of such models imply that model outputs of interest even when aggregated from several runs may not necessarily meet a single best-fit output criterion (Thiele et al., 2014).

Using the categorical calibration method involved defining the calibration criteria, comprising a range of acceptable values for critical model outputs and exploring the model's parameter space to determine the best-fit parameter values based on the calibration criteria. Outputs of interests from the model were grouped into four main themes namely; residential location, job location, home-work distribution among TAZs and mode choice outputs**.** Under each of the four themes, the criteria for specific outputs were set based on estimates derived from the observational data.

Using Netlogo's BehaviorSpace tool, the model parameters were first dimensioned with estimates obtained from the empirical data. With these initial 'empirical-parameter' value settings, the model was run 45 times. The aggregated simulation outputs were then compared to the calibration criteria to ascertain whether the parameter settings as well as the decision-rules implemented by the agents produced the desired outputs. A summary of outputs from the initial simulation compared with the calibration criteria is presented in table 8.3.

Table 8.3: Criteria for categorical calibration of the model and initial simulation outputs[35]

Sub-component	Outputs of interest	Expected Output range	Simulated outputs based on empirical-parameter values
Residential location	Households occupying compound housing	40-54%	65%*
	Households occupying detached housing	5-10%	6%
	Households occupying semi-detached housing	9-15 %	2% *
	Households occupying flat	30-40 %	4% *
	Renting households	40-53 %	68% *
	Owner-occupier households	20-30 %	5% *
	Rent-free households	20-30 %	27%
	Historical-core zone of residence	25-30 %	46% *
	Inner-suburban zone of residence	40-50 %	29% *
	Outer-suburban zone of residence	30-40 %	25%*
	Average distance of home to shopping/market	100-2084m	553.93
	Average distance of home to school (households with children)	100-853m	1493.54
	Average distance of home to transport terminal	100-3432m	1648.08
	Average distance of home to major roads	100-554m	404.62
Job location	Total non-home-based employment locations	60-70%	45%*
	Total home-based employment locations	30-40%	55%*
	CBD employment locations	40-50%	42%
	Zone2 employment locations	7-12%	10%
	Zone3 employment locations	30-40%	34%
	Zone4 employment locations	2-5%	4%
	Zone5 employment locations	5-7%	11%
	Average home-employment location distance (non-home-based jobs)	100-5000m	4621.02

*Initial simulated output values that do not match the expected output range set as the calibration criteria

[35] The upper limits of distance criteria for proximity to schools, proximity to road and proximity to terminals were derived from the mean distance to these facilities calculated in GIS at 35 x 35-meter resolution for the entire metropolis.

Table 8.3 continued: Criteria for categorical calibration of the model and initial simulation outputs

Sub-model	**Outputs of interest**	**Expected Output range**	**Simulated outputs based on empirical-parameter values**
Home-work distribution (TAZs)	Total work-trip origin TAZ-301	2-5%	2%
	Total work-trip origin TAZ-302	10-15%	26%
	Total work-trip origin TAZ-303	15-25%	18%
	Total work-trip origin TAZ-304	25-30%	25%
	Total work-trip origin TAZ-305	20-25%	18%
	Total work-trip origin TAZ-306	11-15%	11%
	Total work-trip destination TAZ-301 + TAZ-302	30-35%	31%
	Total work-trip destination TAZ-303	10-20%	18%
	Total work-trip destination TAZ-304	20-30%	29%
	Total work-trip destination TAZ-305	12-20%	16%
	Total work-trip destination TAZ-306	5-10%	6%
Mode choice	Motorized transport use	60-70%	45%*
	Non-motorized transport use (walking)	30-40%	55% *
	Private-car ownership and use	15-20%	15%
	Public transport	80-90%	85%
	Public transport (Trotro/mini-bus) use	80-90%	89%
	Public transport (taxi) use	10-15%	11%

*Initial simulated output values that do not match the expected output range set as the calibration criteria

For outputs of interests grouped under the emergent residential location patterns, it was found that simulated results for dwelling type and tenancy type distribution as well as residential distribution among the three urban-zones did not fall within the established calibration criteria. In terms of dwelling type distribution, most of the households (65%), far more than expected ended up in compound housing. For tenancy distribution, whereas more of the simulated households (68%) than expected ended up in the rental sector, the model produced fewer than expected (5%) households in the owner-occupier sector. Moreover, under the simulated job location outcomes, the initial model results showed that more workers than expected (55%) had home-based job locations. Consequently, the simulated non-motorized work travel mode choice also did not produce results within the expected output criteria.

These discrepancies were the result of a combination of poorly fitted parameters and the encoded rules implemented by the agents in the model. Details of changes made to behavioural rules will be discussed later in chapter eight under the model verification section. Where the discrepancy related directly to the parameter values, these values were modified to obtain best-fit values as explained below

Choosing the right parameter settings in subsequent stages of the calibration experiment involved a systematic process of using Netlogo's BehaviorSpace tool to explore all possible values of the parameters of interests and their combinations. The initial empirically determined parameter values provided a useful starting point in determining whether the best fit model parameter values would lie either below or above the "empirical-values". Using the full factorial approach, each of the parameters of interest were set to a minimum possible value and then increased systematically in selected intervals up to the maximum possible values. The model was then run repeatedly for all combinations of the parameter values. The simulated outputs for each combination of the parameters were then analysed and compared with the calibration criteria. The combination of model parameters that produced simulated results that matched the calibration criteria were selected as the best-fit parameter values. Table 8.4 shows the model parameter values for the initial empirically determined values and the best-fit parameter values derived through the parameter-sweeping experiments.

Table 8.4: Best-fit parameter values obtained through the model parameter sweeping experiments

Choice Process	Parameter Name	Parameter Value Settings: Empirical-values[36]	Parameter Value Settings: Best-fit-values[37]
Residential choice: Urban zone choice probabilities ***(Range: 0-1)***	Core/suburban choice probability: urban-poor	0.38	0.28
	Inner-suburban /outer- suburban choice probability: urban-poor	0.65	0.65
	Core/suburban choice probability: low-income	0.38	0.28
	Inner-suburban /outer- suburban choice probability: low-income	0.65	0.65
	Core/suburban choice probability: lower-middle-income	0.29	0.19
	Inner-suburban /outer- suburban choice probability: lower-middle-income	0.55	0.55
	Core/suburban choice probability: upper-middle-income	0.22	0.12
	Inner-suburban /outer- suburban choice probability: upper-middle-income	0.51	0.55
	Core/suburban choice probability: high-income	0.17	0.10
	Inner-suburban /outer- suburban choice probability: high-income	0.63	0.63
	Core/suburban choice probability: rich	0.17	0.10
	Inner-suburban /outer- suburban choice probability: rich	0.63	0.63
Residential choice: WTP for housing in rental market (% of monthly income) ***(Range: 0-100)***	WTP Low-income (lower-half)	11%	11%
	WTP Low-income (upper-half)	15%	15%
	WTP Lower-middle income (lower-half)	10%	10%
	WTP Lower-middle income (upper-half)	12%	12%
	WTP Upper-middle income (lower half)	32%	32%
	WTP Upper-middle income (upper-half)	55%	55%
	WTP High income (lower-half)	20%	20%
	WTP High income (upper-half)	31%	31%
	WTP Rich (lower-half)	50%	50%
	WTP Rich (upper-half)	60%	60%
Job choice: Home-based employment choice probability ***(Range: 0-1)***	Home-based employment choice probability: low-skilled at historical-core zone	0.31	0.10
	Home-base employment choice probability: low-skilled at inner-suburban zone	0.23	0.05
	Home-base employment choice probability: low-skilled at outer-suburban zone	0.30	0.05
	Home-base employment choice probability: intermediate-skilled at historical-core zone	0.38	0.05
	Home-base employment choice probability: intermediate-skilled at inner-suburban zone	0.25	0.02
	Home-base employment choice probability: intermediate-skilled at outer-suburban zone	0.31	0.02

[36] Empirical-values of parameter settings are values obtained from the cross-sectional survey data obtained from the Kumasi metropolis

[37] Best-fit values are adjusted parameter values obtained through the systematic exploration of the model's parameter space using Netlogo's BehaviorSpace experiment.

Table 8.4: continued: Best-fit parameter values obtained through the model parameter sweeping experiments

Choice Process	Parameter Name	Parameter Value	
		Empirical-values	**Best-fit-values**
Job choice: Home-based employment choice probability ***(Range: 0-1)***	Home-base employment choice probability: high-skilled at historical-core zone	0.32	0.02
	Home-base employment choice probability: high-skilled at inner-suburban zone	0.23	0.01
	Home-base employment choice probability: high-skilled at outer-suburban zone	0.26	0.01
Travel choice: Motorized travel mode choice (private car vs public transport) ***(Range: 0-1)***	Private car ownership probability: urban poor	0.01	0.01
	Private car ownership probability: low-income	0.01	0.01
	Private car ownership probability: lower-middle-income	0.08	0.08
	Private car ownership probability: upper-middle income	0.25	0.25
	Private car ownership probability: high income	0.57	0.57
	Private car ownership probability: rich	0.78	0.78
Travel choice: Public transport mode choice (mini-bus/Trotro vs taxi) ***(Range: 0-1)***	Mini-bus/Trotro choice probability: urban poor	0.91	0.91
	Mini-bus/Trotro choice probability: low-income	0.88	0.88
	Mini-bus/Trotro choice probability: lower-middle-income	0.91	0.91
	Mini-bus/Trotro choice probability: upper-middle income	0.88	0.88
	Mini-bus/Trotro choice probability: high income	0.79	0.79
	Mini-bus/Trotro choice probability: rich	0.64	0.64

8.3 Final model simulation and analysis of results

A total of 15 simulation iterations constitute a full run of the calibrated model. This means that the model is run for 15 simulation years, starting with the year 2000 as the base year and 2015—the year for which the observational data used for the model calibration was collected—as the final year. The initial land price distribution is based on available data in 2000. Due to the absence of fine grain data on household characteristics dating back to the year 2000, the socio-demographic profiles of households derived from the survey data were assumed for the starting household population. Indeed, comparing aggregate trends based on available census data for 2000 and 2010 showed that in proportionate terms, family structure and composition as well as life-cycle stage characteristics such as coupling rate in the case study metropolis have remained stable over the period.

As explained in chapter seven, section 7. 4, aspatial data used for the model implementation was at the level of households and individual workers within the households. The entire metropolis was divided into a 30.48m x 30.48m lattice. The size of the world was 201 x 201, comprising a total of 13, 272 cells of Netlogo patch size 4, representing land parcels". The metropolis also had two main sub-divisions namely; the three urban-zones and the six macro TAZ system. The basic spatial unit within the model is therefore the 30.48m x 30.48m cell. The location of dwellings and households' place of residence were simulated at this fine scale. In the presentation of the results however, the urban-zones and TAZ system constitute the different spatial scales of aggregation of the model outputs. For example, while residential location patterns of households are aggregated at the level of urban-zones (i.e. historical-core, inner-suburb and outer-suburb), O-D patterns derived from the home-work location pairs of the workers within the household are aggregated at the level of TAZs.

To consider the effect of stochasticity on the outputs of interest, each iteration was repeated 45 times in line with the minimum model run estimated earlier using the co-efficient of variation metric. The final results presented for each year is therefore an aggregation of the outputs of the simulation across these multiple runs.

Furthermore, as the simulation results will show, the current model run initiates with a relatively small sample population (i.e. 30, 000 households). Thus, one of the goals of the simulation is to see whether by starting with a relatively small sample population and implementing the endogenous population growth mechanism, leading to the formation of new households in subsequent iterations, one can generate realistic patterns at scale of the case study metropolis.

In the sections that follow, results of the simulation are discussed.

8.3.1 Background Characteristics of simulated household population

As shown in table 8.5, at the beginning of the simulation, a total of 30,000 households were initialized. Because of the formation of new households, the total generated household population increased to 63608 at the end of the simulation. Thus, on the average, new households are formed endogenously from the starting population at a rate of 5.14%. In terms of the actual number of households that got settled in the metropolis (i.e. became successful in their residential location search), the results of the simulation show that a total of 58,217 households, representing 92% of the total household population generated, was achieved at the end of simulation.

Table 8.5: Total simulated households across the model simulation.

	Total generated households		**Total settled households**		
Iterations	number	Std. deviation	number	Std. deviation	**Percentage settled**
0	30000	-	0.00	0.000	0
1	52219	102.60	20914	1424.30	40
2	52924	99.77	32130	1700.87	61
3	53601	102.63	38411	1927.12	72
4	54255	104.01	42377	1997.47	78
5	54879	101.19	45065	1859.14	82
6	55484	102.87	46937	1774.99	85
7	56014	176.85	48303	1691.74	86
8	56571	190.57	49456	1668.42	87
9	57112	258.15	50470	1669.38	88
10	57650	424.57	51187	1991.07	89
11	58233	777.67	52188	1848.12	90
12	58954	1529.99	53132	2315.05	90
13	59795	2991.05	54138	3507.51	91
14	61162	5944.62	55623	6293.67	91
15	63608	11874.39	58217	12108.68	92

Figure 8.1 depicts the residential household population and growth rate trend over time. On the average, new households settle in the metropolis at an annual rate of 7.06%. The trend shows that

the rate of settlement increases sharply initially between the 1st and 2nd iterations at a rate of 2.9% but falls and stabilizes across the subsequent iterations.

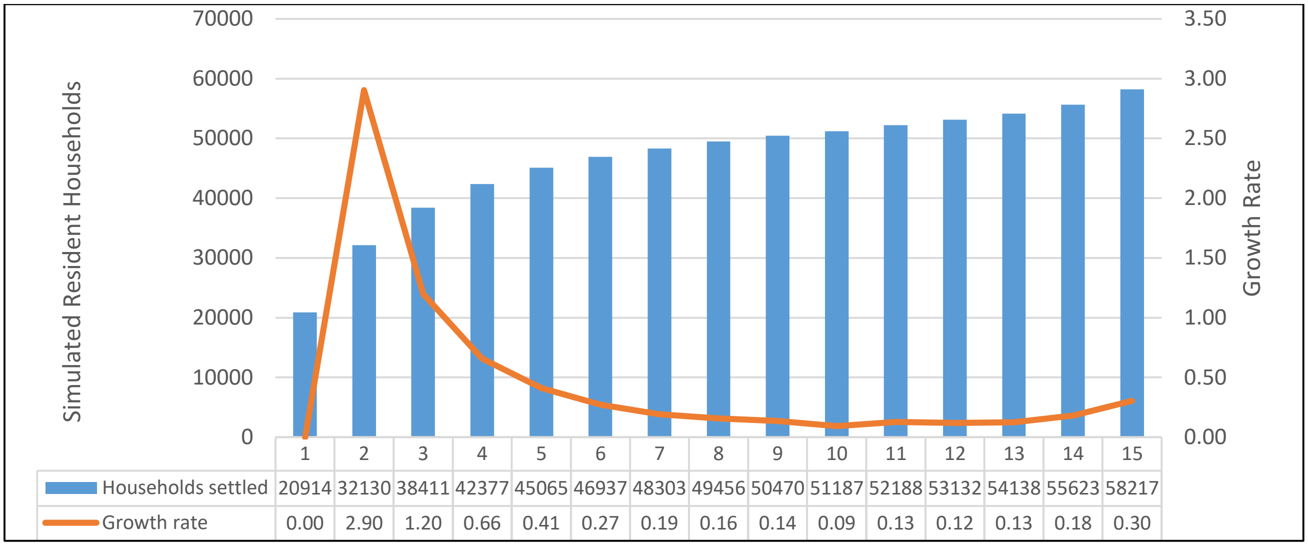


Figure 8.1: Simulated residential household population size and growth rate across the model iterations

The key distinguishing characteristics of households as indicated under the socio-demographic sub-component of the model were income, life-cycle stage and household size and composition. Figure 8.2 shows the proportion of the total residential household population under each of the five income categories defined.

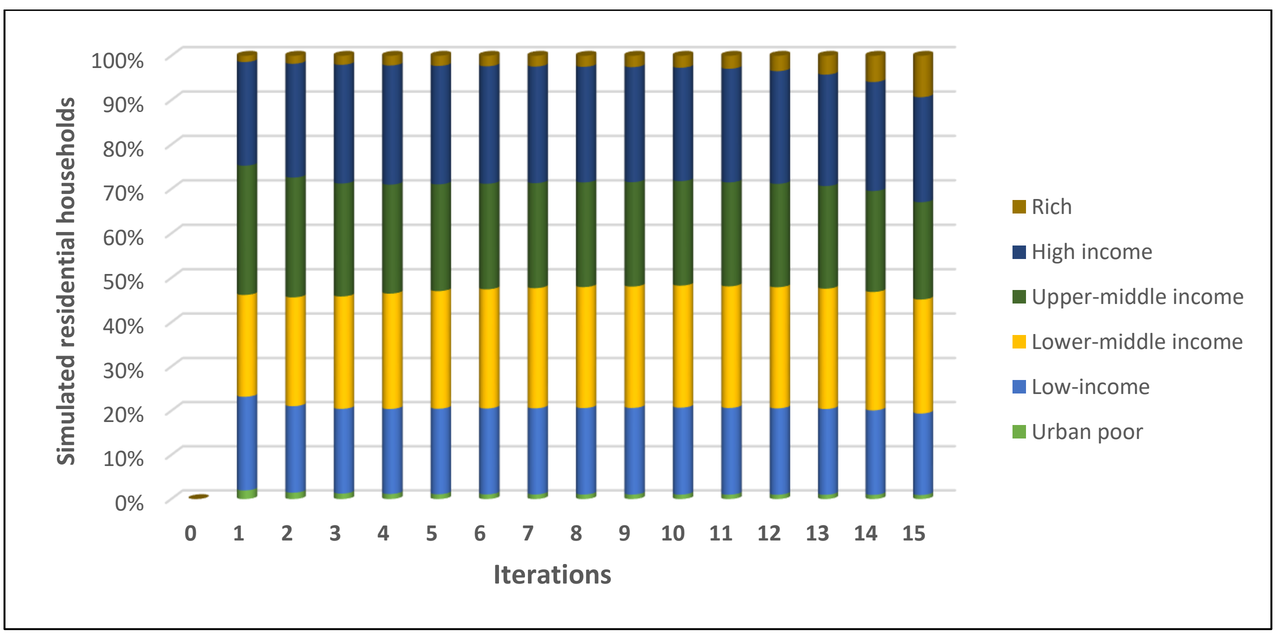


Figure 8.2: Distribution of simulated residential households among income groups across the model iterations

Among the total residential households generated across the model simulation, urban poor and low-income households constituted one percent and 20% respectively. Lower-middle and upper-middle income households constituted 26% and 24% of the simulated household population respectively while 26 % and three percent of households fell into high income and rich categories respectively.

Moreover, under life-cycle-stage and family characteristics, the model results show a coupling rate of 80%; cohabiting, single and divorced households constitute one percent, 15% and two percent of households respectively. About, 88% of all households at the end of the simulation had children. Figure 8.3 shows the percent distribution of households with children across the six income categories at the end of the simulation.

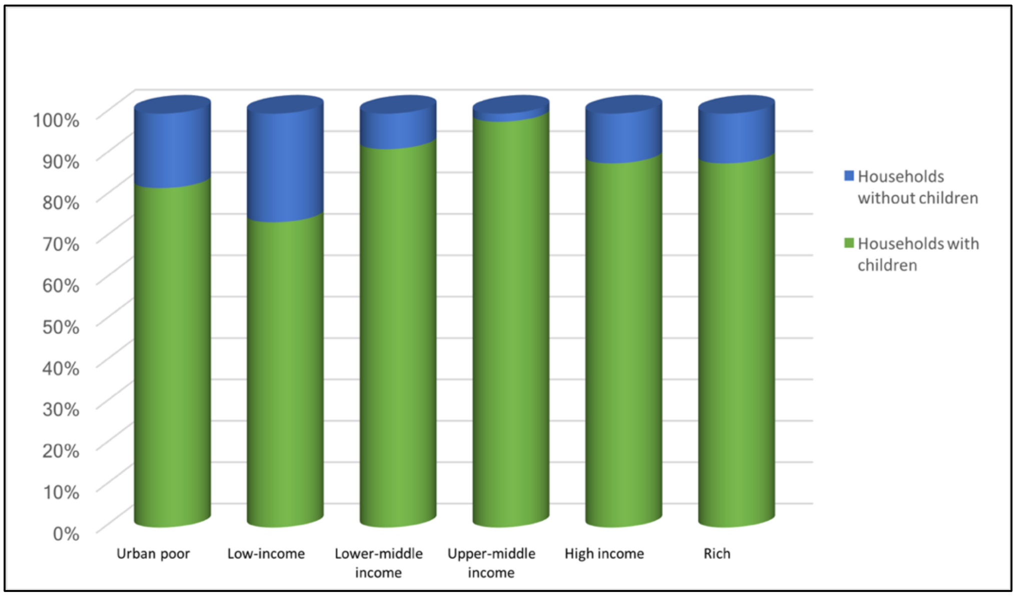


Figure 8.3: Proportion of simulated residential households across income groups with and without children

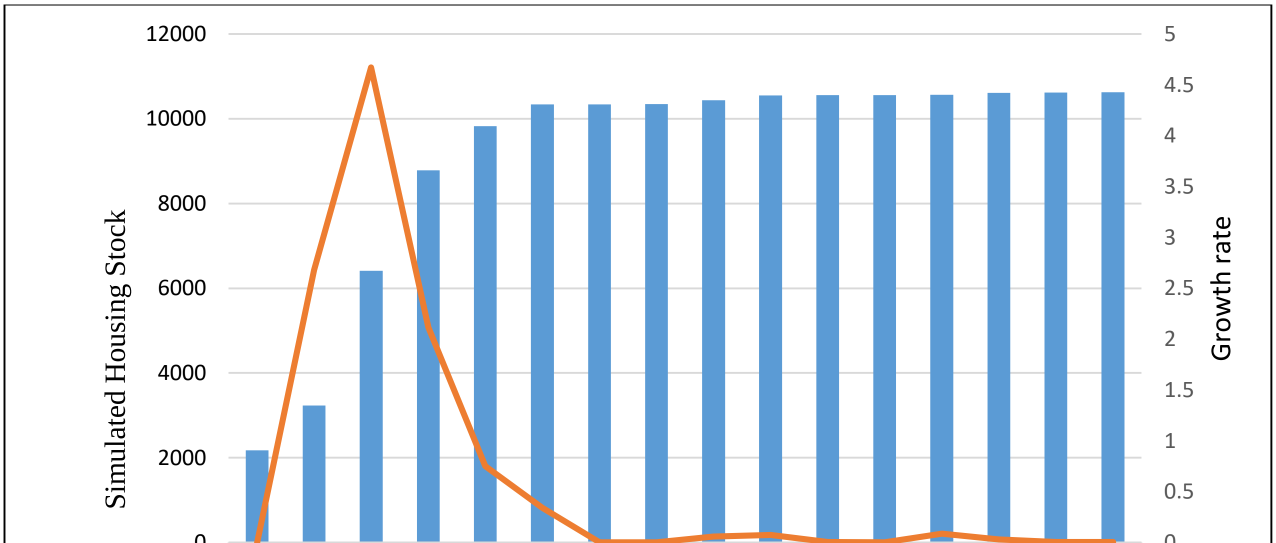
8.3.2 Simulated residential location patterns

The emergent residential location patterns of the model are examined in terms of the total dwellings developed in the metropolis as well as the characteristics of locations realized by the households in terms of dwelling type and tenancy arrangements obtained, and proximity to essential amenities and infrastructure based on their preferences. These are discussed in the sections that follow.

Emergent distributions of dwellings and dwelling characteristics

The total number of dwellings developed in the metropolis across the model simulation results directly from the interplay between population size and existing vacant dwelling on each iteration. The feedback relationship between demand and supply, which was captured as either a housing deficit or surplus, thus, determine the additional dwelling units generated in space and time.

Figure 8.4 shows the housing supply trends across the model iterations. From an initial dwelling units of 2,175 at the beginning of the simulation, the endogenous demand-supply feedback mechanism in the model generated a total of 10, 625 dwelling units at the end of the simulation. Thus, on the average, new dwellings were generated at a rate of 8.25%. As depicted in figure 7.27, the rate of supply of new dwellings is not uniform across time. From a rate of 2.68% on the 1st iteration, the housing supply rate increased to 4.6% in the 2nd iteration. By the 3rd iteration, the rate of supply decreased to 2.11%, beyond which it stabilized. What this means is that although new houses continue to be added to the old stock in response to population increase over time, the prevailing vacancy in the old stock is taken into account. This ensures that the cumulative increase in dwellings is consistent with the household population actively searching for residential locations in the property market.



Housing Stock	2175	3234	6414	8782	9827	10341	10341	10342	10436	10553	10558	10559	10562	10611	10618	10625
Growth Rate	0	2.6802	4.6703	2.1174	0.7519	0.3406	0.001	0.0011	0.06	0.0741	0.0062	0.0009	0.087	0.0308	0.0043	0.0043

Figure 8.4: Distribution of housing units in the metropolis across the model iteration

A breakdown of the distribution of housing stock among the three broad zones—historical-core, inner-suburb and outer-suburb of the metropolis in space and time is depicted in figure 8.5 whiles figure 8.6 shows images captured from the model interface showing the simulated location of dwelling units in the metropolis over time. Consistent with the existing distribution of dwellings in the Kumasi metropolis, most of the dwellings generated at the end of the simulation were in the outer-suburban zone (53%) followed by the historical-core (25%) and the inner-suburban zone (22%).

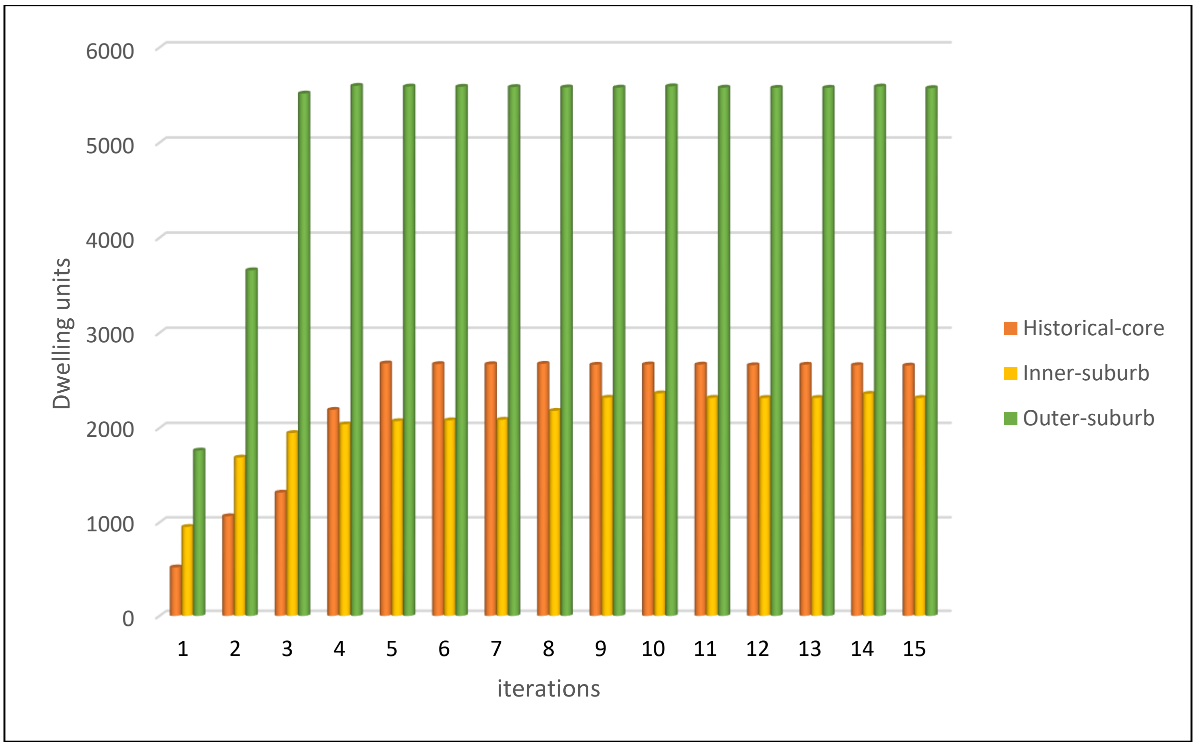


Figure 8.5: Distribution of housing units among the three urban-zones across the model iteration

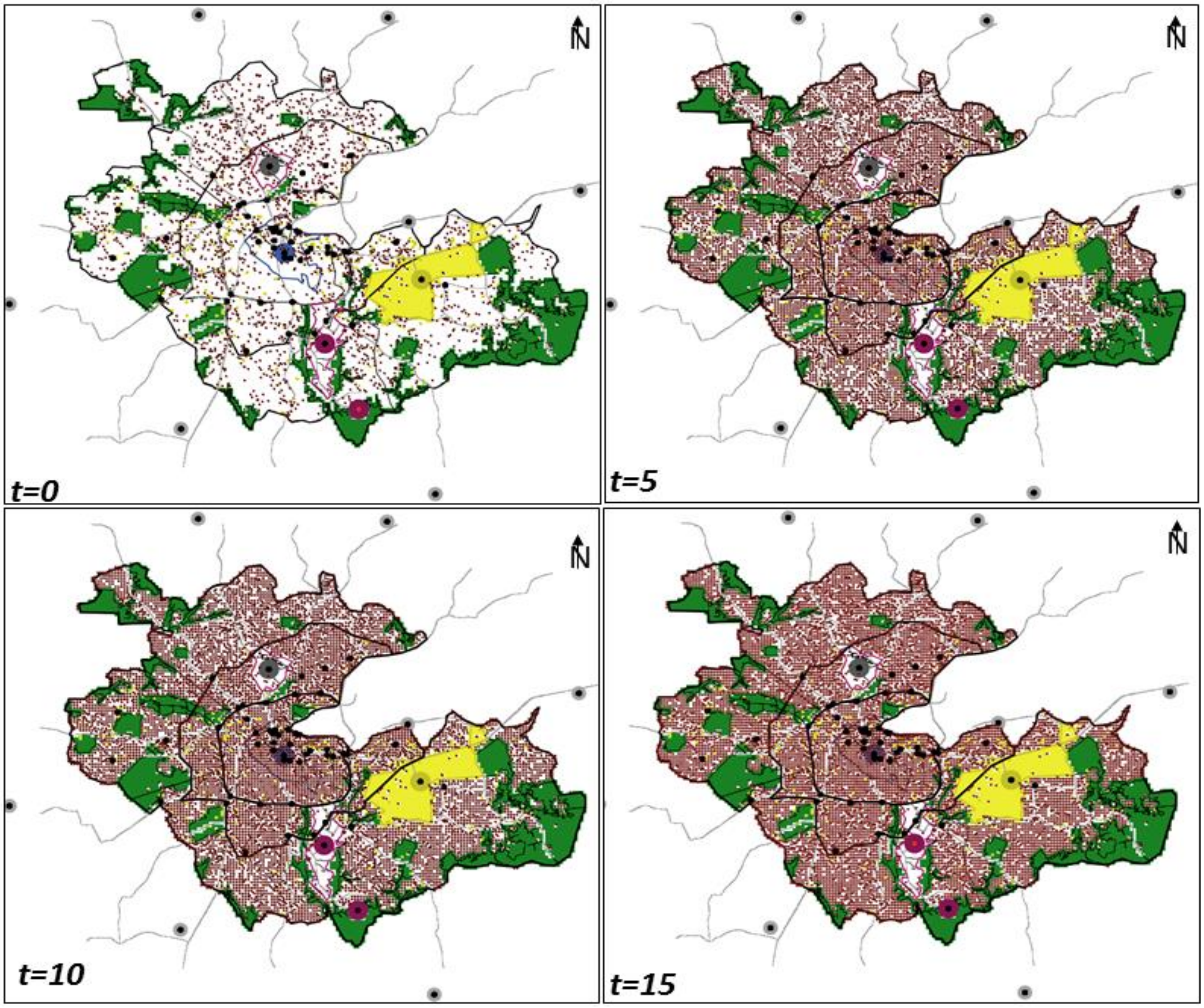


Figure 8.6: Spatio-temporal distribution of the location of household's dwellings. Note: Red points indicate dwellings

The number of dwellings found in each zone is a function of the number of household population actively looking for residence there, subject to the availability of vacant plots required to accommodate the new dwellings generated in response to the prevailing demand as well as the level of expansion or intensification permitted. Consequently, the outer-suburban zone, being the largest of the three urban-zones provides more land for housing to accommodate demand generated locally within it and globally from the other two urban zones to mimic the process of outward expansion. The historical-core on the other hand, being the preferred location for households of lower incomes who constitute the largest share of the population, responds to demand, by allowing intensification, where new dwellings could be accommodated on existing built parcels, mimicking the process of dwelling extensions and redevelopment. This mechanism is reflected in the 4th iteration onwards where despite the historical-core being fully built, continues to receive more of the new dwellings compared to the outer-suburban zone (see figure 8.6).

Furthermore, the analysis examined the distribution of the simulated dwellings by type among the various urban locations. As shown in figure 8.7, compound housing constituted the dominant housing type overall (52%) and across the three urban-zones. However, the proportion of compound dwellings remains larger in the historical-core (60%) and decreases at the inner-suburban (50%) and outer-suburban (48%) zones.

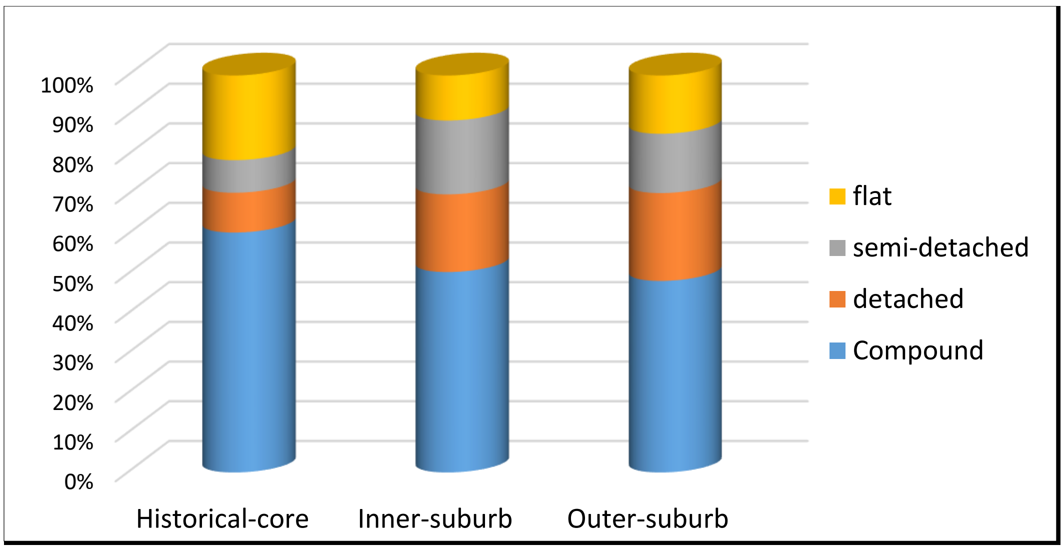


Figure 8.7: Distribution of simulated dwelling types in the three urban zones of the metropolis at the end of the simulation

On the contrary, the proportion of detached housing remains significantly higher in the inner-suburban (20%) and outer-suburban (22%) zones compared to the historical-core (10%). Similarly, whereas 19% and 15% of all dwellings generated by the simulation were semi-detached dwellings located in the inner and outer suburban zones respectively, a relatively small percentage of semi-detached dwellings (eight percent) as expected, was generated in the historical-core. Thus, using the dwelling type distributions from the case study area at initialization, the model was well able to generate new dwellings of types consistent with what pertains at the metropolitan scale and among the three urban zones.

Emergent residential locations realized by the simulated households

Patterns of residential locations achieved by the households, based on their preferences and income levels, are analysed in terms of the distribution of the simulated residential locations of the households among the three urban zones and the dwelling types and tenancies realized by households of the different income-groupings represented in the model. Table 8.6 provides a summary of the location of households' homes among the three urban-zones across the model iterations. Overall, most of the households (41%) found residential locations in the inner-suburban locations of the metropolis compared to those in the historical-core (28%) and outer-suburb zone (31%).

Table 8.6: Residential locations realized by the simulated households at the level of urban-zones

Iterations	**Simulated households' distributions**				**Percentages**		
	Historical-core	Inner-suburb	Outer-suburb	Total	Historical-core	Inner-suburb	Outer-suburb
1	4242	9052	7620	20914	20	43	36
2	6575	14115	11441	32130	20	44	36
3	7953	17047	13410	38411	21	44	35
4	8948	18842	14587	42377	21	44	34
5	9765	19987	15314	45065	22	44	34
6	10302	21013	15623	46937	22	45	33
7	10761	21613	15929	48303	22	45	33
8	11088	21969	16399	49456	22	44	33
9	11450	22384	16636	50470	23	44	33
10	11674	22615	16899	51187	23	44	33
11	12073	23059	17056	52188	23	44	33
12	12557	23288	17287	53132	24	44	33
13	13176	23581	17380	54138	24	44	32
14	14403	23714	17506	55623	26	43	31
15	16559	23863	17796	58217	28	41	31

As encoded in rules of behaviour implemented by the households in their residential location search, preference for a particularly urban-zone is determined by the income category of the households involved. In view of this, the analysis examined the income categories of the households in each of the urban-zones over time. For the purpose of this analysis, the income categories of the households have been reduced to three groups by merging urban-poor and low-income households into 'low-income', lower-middle and upper-middle household into 'middle-income' and high and rich households into 'high-income'. The results are depicted in figure 8.6.

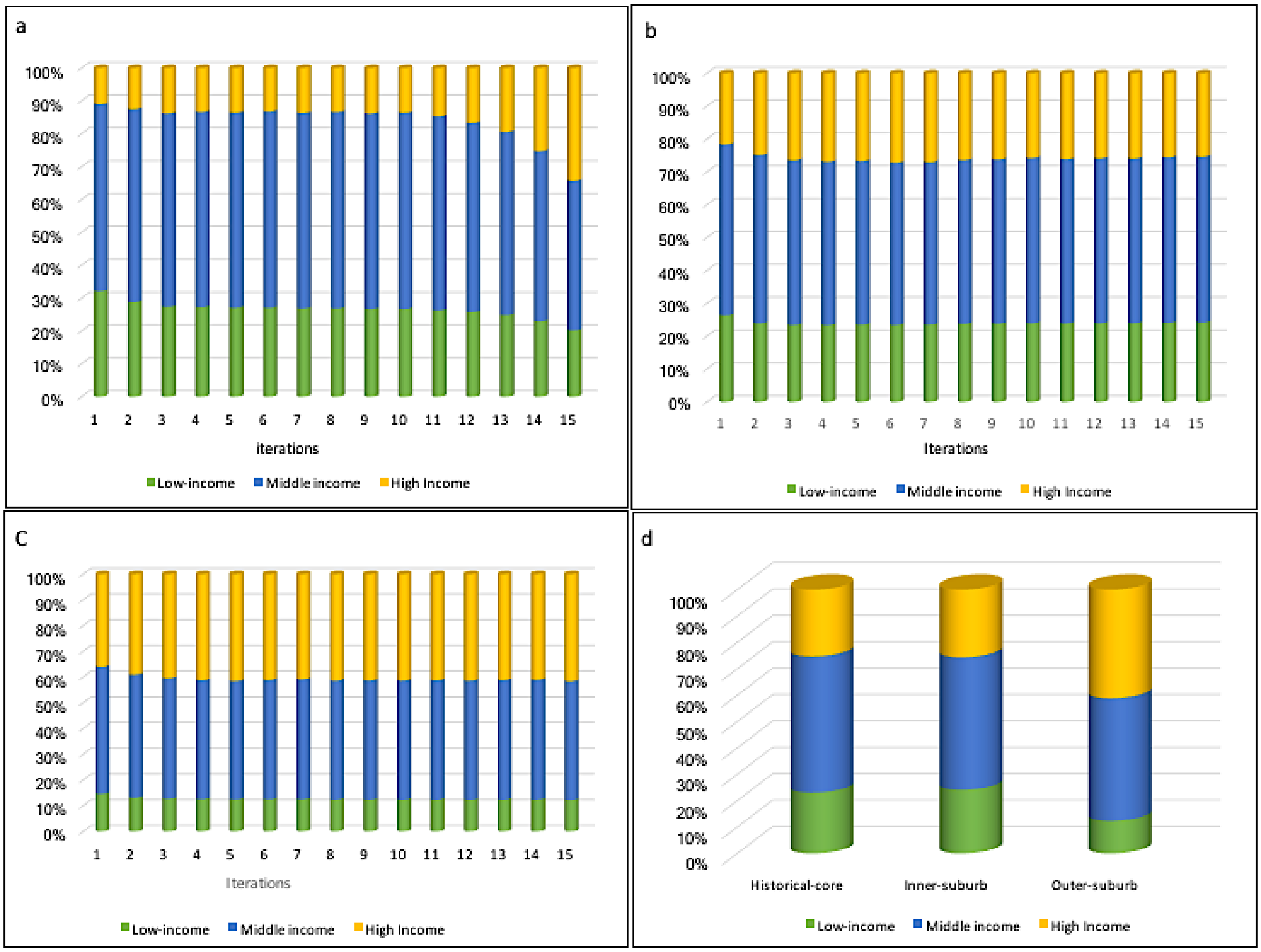


Figure 8.8: Distribution of settled households by income grouping in (a) historical-core (b)inner-suburb (c) outer-suburb (d) urban-zones on the 15th iteration

Overall, the simulation generated outputs consistent with the observed residential location distributions in the case study metropolis. As shown in figure 8.8, the proportion of low-income households having their residence at the end of the model simulation in the historical-core and inner-suburb were significantly higher at 23% and 24 % respectively compared to the 12% living in the outer-suburb. Middle income households found residence across all the three urban zones although the proportion is higher in the historical-core (i.e. 53%) but decreases to 50% and 47% in the inner and outer suburban-zones respectively. Among high income households however, the majority (41%) achieved residential locations in the outer-suburban zone while 25% and 26% realized their home location preferences in the historical-core and outer-suburban zones respectively.

In each of the three urban-zones, the simulated households also realized preferred attributes of dwellings subject to their income levels and competition with other households with similar preferences. Table 8.7 summarizes the dwelling types occupied by the simulated households. On the final iteration of the model, a larger proportion of households realized their residential locations in compound housing (43%) and flats in multi-storey buildings (40%). Consistent with the observed dwelling types occupied by households in the case study metropolis, relatively fewer households found accommodation in detached (eight percent) and semi-detached (nine percent) dwellings.

Table 8.7: Dwelling types realized by households at their residential locations

Iterations	**Simulated households occupied dwelling types**					**Percentages**			
	Compound	Detached	Semi-detached	Flat	Total	Compound	Detached	Semi-detached	Flat
1	8800	1982	1866	8277	20925	42	9	9	40
2	13460	2885	2932	12885	32162	42	9	9	40
3	16018	3414	3604	15408	38445	42	9	9	40
4	17764	3758	3993	16894	42410	42	9	9	40
5	18915	3980	4293	17908	45096	42	9	10	40
6	19806	4136	4513	18510	46965	42	9	10	39
7	20326	4400	4659	18943	48328	42	9	10	39
8	20897	4502	4767	19313	49479	42	9	10	39
9	21357	4452	4991	19691	50491	42	9	10	39
10	21792	4534	5077	19803	51207	43	9	10	39
11	22121	4589	5141	20355	52206	42	9	10	39
12	22543	4637	5067	20900	53148	42	9	10	39
13	22805	4709	5136	21502	54153	42	9	9	40
14	23080	4741	5189	22628	55637	41	9	9	41
15	24954	4781	5216	23266	58217	43	8	9	40

Three important dimensions of the simulated results, in terms of the distribution of the achieved dwelling types of the households across the three main income categories and within the three urban-zones are presented in table 8.8 and table 8.9 respectively.

Table 8.8: Simulated dwelling types occupied by households of different income categories in the metropolis

	Simulated dwelling types by income groups				**Percentages**		
	Low-income	Middle-income	High-income	Total	Low-income	Middle-income	High-income
Compound	6680	15779	2495	24954	27	63	10
Flat	1419	9709	1485	12613	11	77	12
Detached	1598	1697	12138	15434	10	11	79
Semi-detached	1763	2019	1434	5216	34	39	27

Table 8.9: Simulated dwelling types occupied by households within the three urban-zones

	Simulated dwelling types in urban-zones				**Percentages**		
	Historical-core	Inner-suburb	Outer-suburb	Total	Historical-core	Inner-suburb	Outer-suburb
Compound	7591	11041	6322	24954	30	44	25
Flat	8547	6076	8643	23266	37	26	37
Detached	586	2597	1598	4781	12	54	33
Semi-detached	72	4352	792	5216	11	50	31

Figure 8.9 combines all the three dimensions of the above results of the simulation. The analysis show that the proportion of low-income and middle-income households successfully realizing their homes in compound housing is significantly higher in all three urban zones compared to high income households (figure 8.9a).

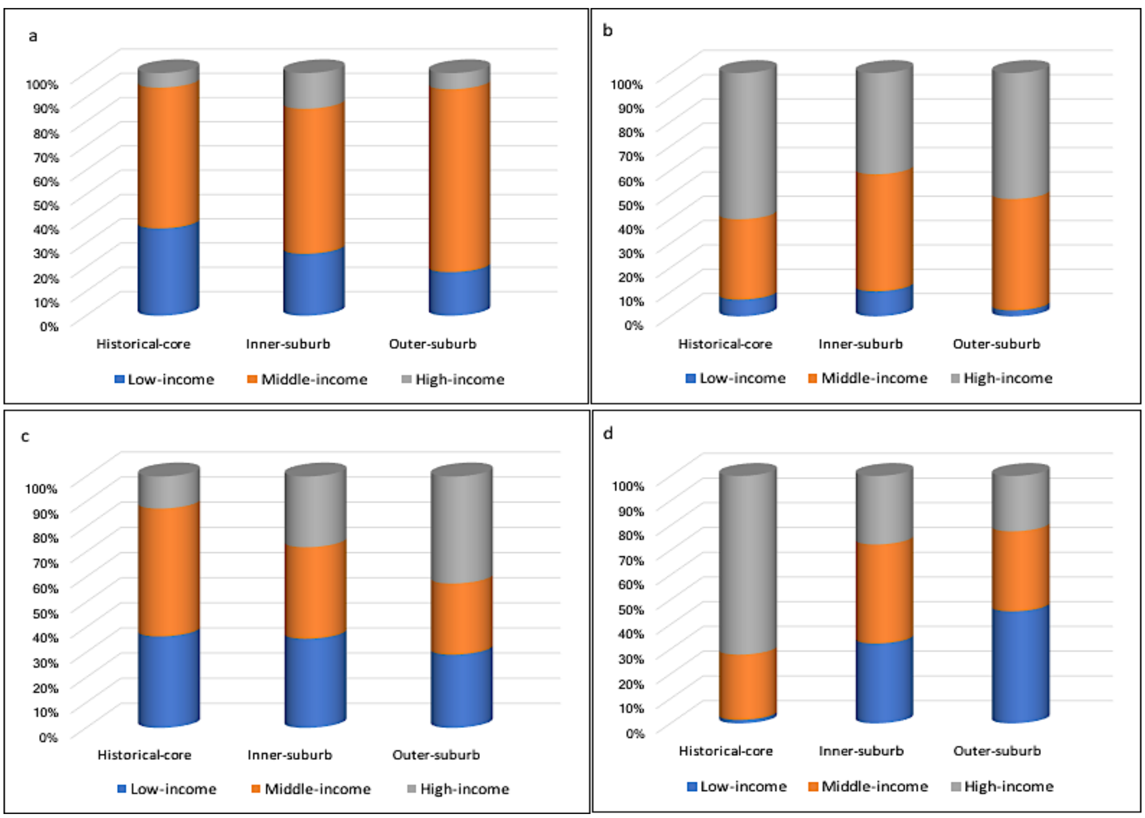


Figure 8.9: The proportion of households of different income groups occupying (a) compound (b) flat (c) detached and (d) semi-detached in the three urban zones

Moreover, within flats in multi-storey buildings (figure 8.9b), middle-income and high-income households are the main occupants, representing 42% and 52% respectively; the proportion of high income households remains significantly higher in the historical-core (60%) inner-suburb (42%) and outer-suburb (52%), followed by middle-income households—historical-core (33%) inner-suburb (48%) and outer-suburb (56%) and low-income households of which seven percent, 10% and six percent occupy flats in historical-core inner-suburban and outer-suburban zone respectively. The distribution of households within semi-detached (figure 8.9c), and detached housing (figure 8.9d) shows a general trend where more middle-income and high-income households realized residence in these dwellings across the three urban-zones compared with low-income households.

An equally important aspect of households' residential locations simulated in the model is the tenancy types achieved by the household agents. Table 8.10 shows the proportion of owner-occupier, renting and rent-free households on each iteration of the simulation. The results show that at the end of the simulation, 40% of all households successful with the residential location search process were renters, 30% became owner-occupiers while the remaining 30% found residence in the rent-free sector.

Table 8.10: Tenancy types realized by households at their residential locations

Iterations	**Simulated Tenancy arrangements**				**Percentages**		
	Renting	Owner-occupier	Rent-free	Total	Renting	Owner-occupier	Rent-free
1	10126	4773	6036	20935	48	23	29
2	14646	8329	9204	32179	46	26	29
3	16807	10521	11135	38462	44	27	29
4	18150	11755	12519	42425	43	28	30
5	19038	12496	13577	45110	42	28	30
6	19637	12968	14372	46977	42	28	31
7	20095	13281	14963	48339	42	27	31
8	20513	13521	15454	49488	41	27	31
9	20840	13783	15877	50500	41	27	31
10	21136	13842	16236	51215	41	27	32
11	21389	14284	16540	52213	41	27	32
12	21612	14729	16814	53154	41	28	32
13	21837	15277	17045	54159	40	28	31
14	22047	16327	17269	55643	40	29	31
15	23178	17584	17455	58217	40	30	30

A summary of the realized tenancy types within the four main dwelling types, which constituted the household's choice alternatives, and the three-broad urban-zones is presented in figure 8.10**.**

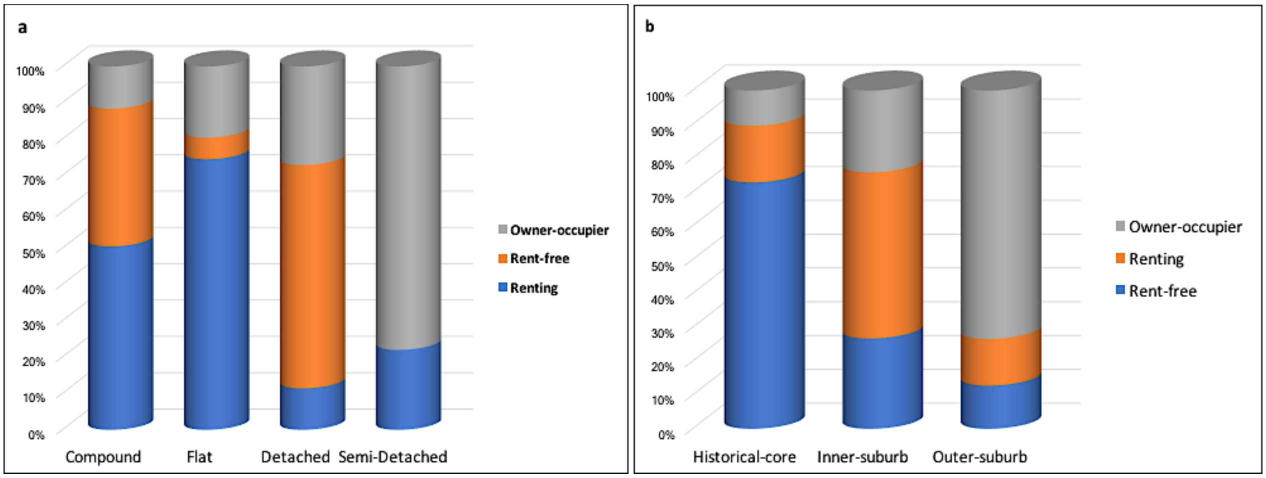


Figure 8.10: The proportions of households' tenancies realized in the (a) different housing types and (b) urban-zones

Furthermore, figure 8.11 shows the distribution of the households achieved tenancies in the different housing types and urban-zones of residence. The results show that living rent-free was prevalent among low-income (45%) and middle-income (55%) households only. About 60% of rent-free tenancy status among the households was found in compound housing: the remaining households in this non-market sector were found in flats (19%) and detached housing (21%). Among the urban zones, most of the households realized their preferred rent-free accommodation in the historical-core (44%) and inner-suburban zone (40%).

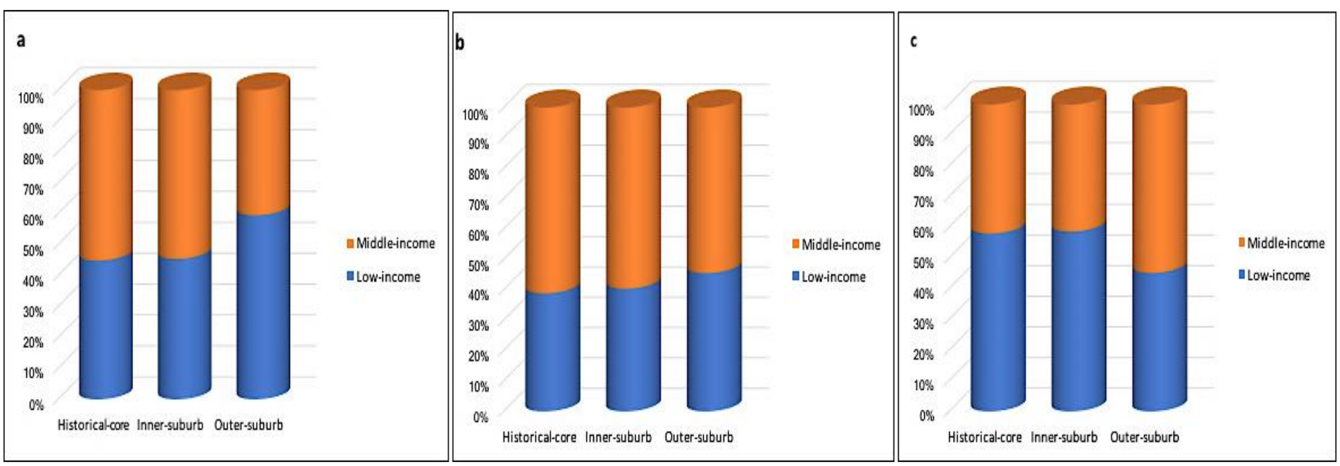


Figure 8.11 Proportions of rent-free households of different income groups in the urban-zones within (a) compound (b) detached (c) flat

A breakdown of the distribution of the households renting their accommodation within the different housing types and urban-zones of residence is shown in figure 8.12.

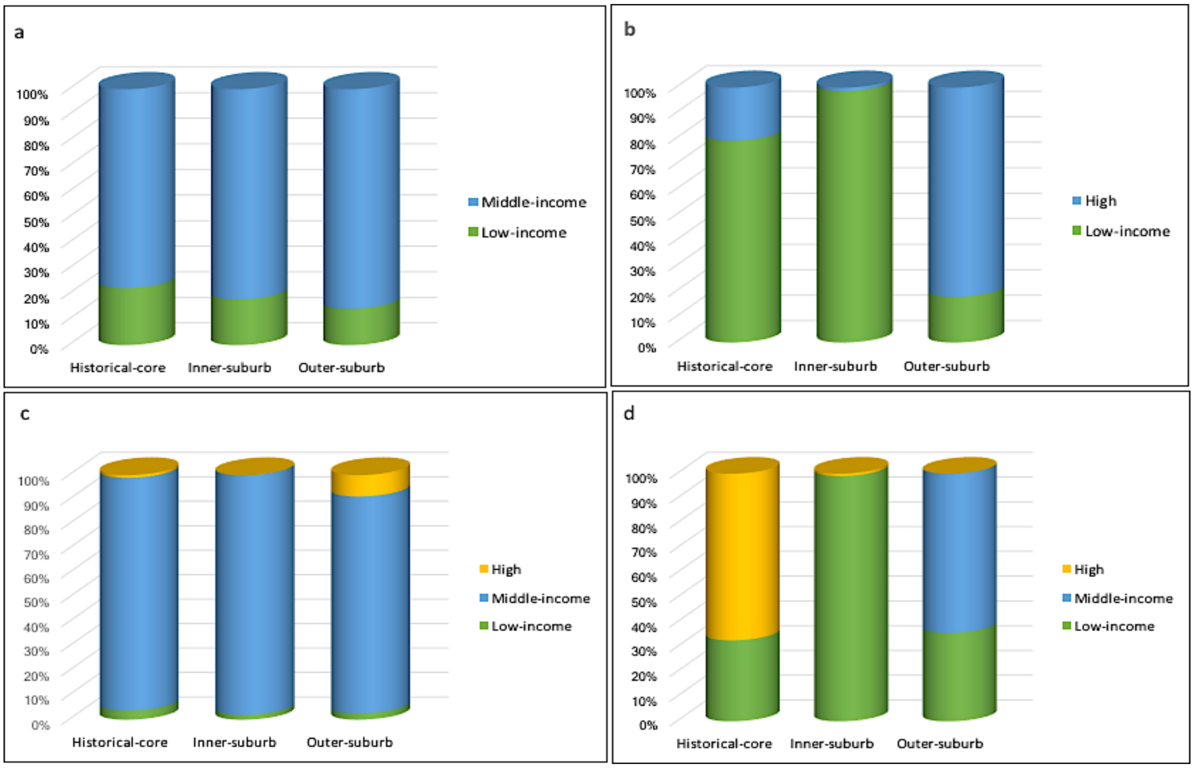


Figure 8.12: Proportion of households of different income groups
Renting (a) compound (b) detached (c) semi-detached (d) flat in the three urban zones

Renting as a tenure choice, cuts across all households, although more middle-income households (82%) realized residence in the rental sector compared to low-income (15%) and high-income household (3%). A greater percentage of rented accommodation was realized by the households in compound housing (25%) and flat (73%); the remaining two percent were found in detached and semi-detached houses. Moreover, at the end of the simulation, 73% of household renting their accommodation did so in the inner-suburban zone while 10% and 27% of rental housing was achieved in the historical-core and outer-suburban zones respectively.

Owner-occupied tenancy was concentrated in semi-detached (63%) housing mainly in the outer-suburban zone (68%). The proportion of owner-occupier households within the dwelling types providing their accommodation and across the three urban zones is shown in figure 8.13.

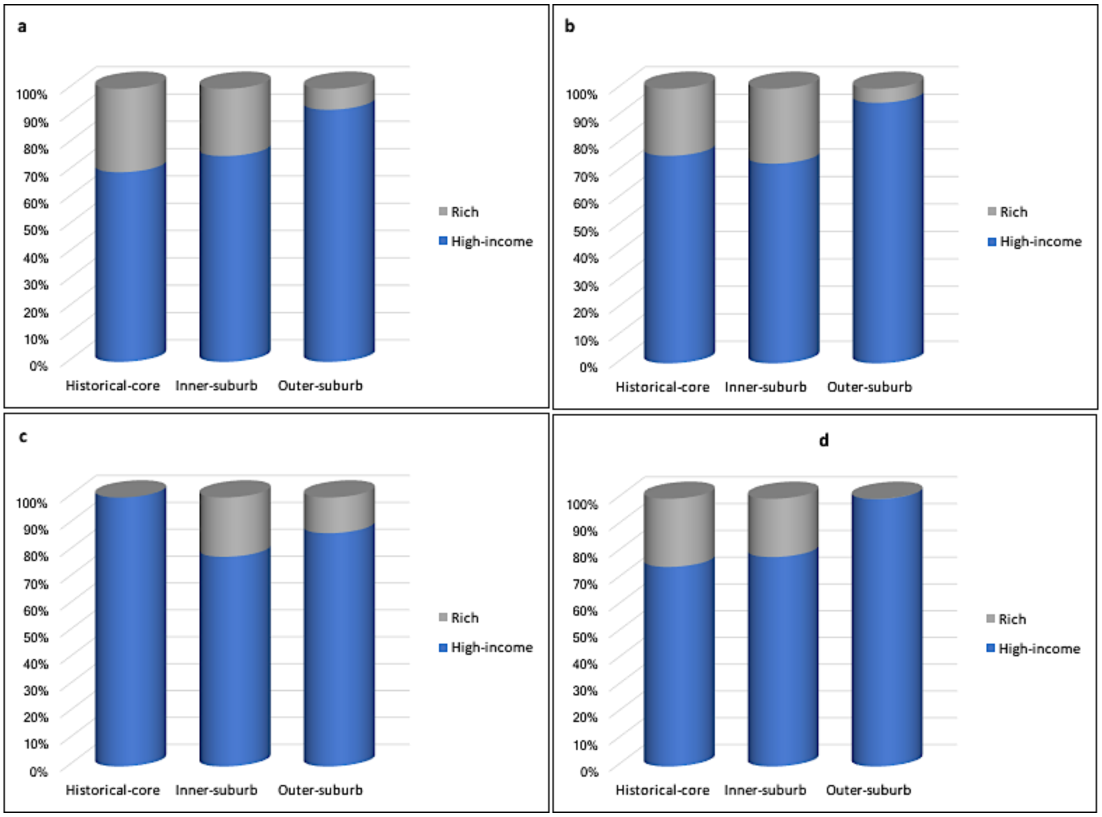


Figure 8.13: Proportions of owner-occupier households' tenancy of different income groups in the urban-zone within (a) compound (b) detached (c) semi-detached (d) flat

8.3.3 Households' home locations proximity to amenities and infrastructure

In addition to preference for urban-zones, dwelling types and residential tenancies, households evaluated their residential locations based on some expected distance separation between their home locations and four key amenities namely; basic schools, local markets/shopping, transport terminals and major roads. Thus, the third important output measure in relation to the simulated residential location patterns in the model is the emergent proximity of the households' home location to these essential amenities and facilities.

Table 8.11 provides a summary of the respective distances to amenities achieved by households of the different income groups at their residential locations. In figure 8.14, the evolution of households attained proximities to the amenities on each iteration of the simulation is presented. Overall, all households with children who therefore valued proximity to basic schools, achieved an average home-school distance of 613m (0.61km). An average distance of 1,558m (1.56km); 1, 821m (1.82km) and 448m (0.45km) to local markets, transport terminals and major roads respectively, was realized at the simulated home locations of all households. Consistent with the encoded preferences, households of lower incomes achieved residence relatively closer to all essential amenities than those of higher socio-economic means.

Table 8.11: Simulated distances of households' residence to amenities

Household types	Distance to Amenities (m)							
	Basic[38] school	Std. deviation	Local markets	Std. deviation	Terminals	Std. deviation	Major roads	Std. deviation
Urban-poor	508.68	30.84	1380.92	48.92	1350.56	46.52	379.79	15.70
Low-income	516.83	8.46	1410.22	11.04	1490.21	24.70	401.19	4.48
Lower-middle income	597.81	8.16	1528.56	14.29	1749.23	25.75	438.08	6.35
Upper-middle-income	636.76	21.74	1593.39	35.54	1969.47	60.81	448.37	11.13
High-income	709.20	268.71	1708.68	393.98	2265.48	696.16	495.36	148.41
Rich	513.14	79.58	1393.24	242.62	1133.15	219.19	403.12	78.81
ALL	**613.63**	**86.18**	**1558.82**	**120.12**	**1821.41**	**223.93**	**448.89**	**45.46**

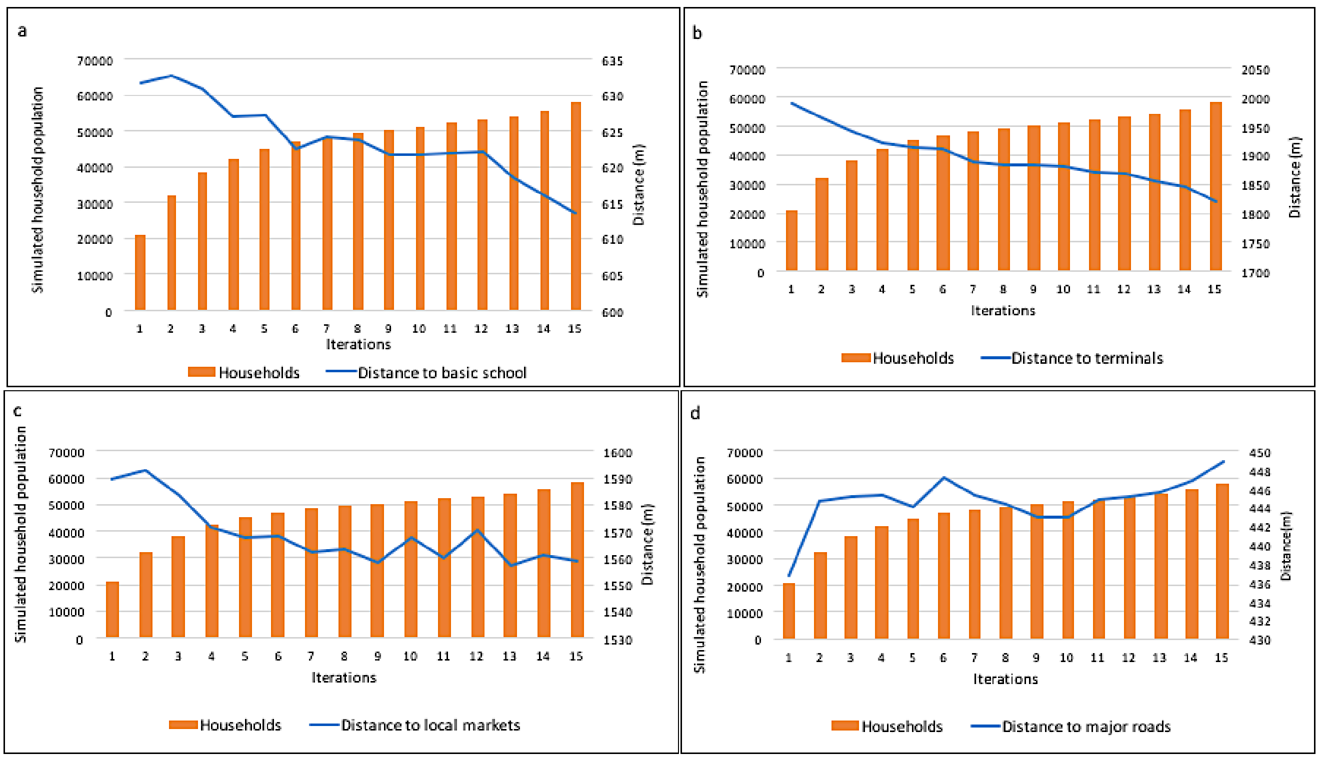


Figure 8.14: Evolution of attained distances to essential amenities at the households' residential locations

The results of the simulation presented in figure 8.14 show that attained distances to basic school, local markets and transport terminals remain relatively higher around the 1st and 3rd iterations of the simulation but falls as population increases over time. The reverse is true with respect to the

[38] Proximity to basic schools was calculated for households with children only.

attained distances to major roads; attained road distances remained relatively lower around 0.43km at the initial stages of the simulation but increase as the number of settled households' increases.

The number and spatial distribution of the amenities could explain the simulated amenity proximity trends. Basic schools, transport terminals and local markets were distributed at various densities throughout the metropolis. Thus, with the existing distribution coupled with the encoded objective to achieve the minimum possible distance separation between the amenities and the place of residence among all household types, the household agents tend to achieve relatively shorter distances to these facilities even as more households settle in the metropolis. The road network being the main arterials, however, traverse the most central locations of the metropolis. Thus, as more households settle in the metropolis and residential locations spread outwards to the suburban zones, attained distances to major roads are expected to rise as the simulated results show. Besides the spatial distribution, it is worth mentioning that the overlaps in expected proximity ranges among the different households, the attributes of dwellings as against the preferred dwelling attributes of the households as well as the households' ability to afford property at any given location, interacted to shape the emergent amenity proximity levels.

8.3.4 Price evolution in the property market

The formation and submission of bid prices based on income levels and willingness to pay on the part of the active household agents, and property prices and the competitive interaction between the purposive household agents, characterized property market transactions in the model. These processes, culminated in the endogenous formation of property prices in the model over time.

The emergent property prices in the rental market at the end of the simulation is presented in table 8.12. The results show house rents for different size dwellings ranging from 1-bedroom to 5-bedrooms. The simulated house rents are also differentiated under each of the four main dwelling types—detached, semi-detached, flat and compound and by the three urban zones—historical-core, inner and outer-suburban zones. Since the results are aggregated across multiple runs, the accompanying standard deviations are also provided. Given that compound houses, although having multiple rooms are rented out by single bedrooms, the results are presented as such.

Table 8.12: Emergent prices in the rental market for dwelling size dwellings at the end of the simulation

House Types	Urban-zone	Simulated house rent (GH¢) by number of bedrooms					
		1	2	3	4	5	>5
Detached	**Historical-core**	**53**	**69**	**140**	**237**	**439**	**648**
	Std.deviation	*12.47*	*27.05*	*42.18*	*53.81*	*21.65*	*21.19*
	Inner-suburb	**91**	**123**	**157**	**253**	**442**	**638**
	Std.deviation	*90.97*	*25.32*	*45.13*	*53.28*	*16.49*	*25.53*
	Outer-suburb	**56**	**128**	**168**	**228**	**452**	**636**
	Std.deviation	*18.30*	*48.04*	*15.17*	*14.68*	*11.46*	*9.35*
Semi-Detached	**Historical-core**	**78**	**116**	**174**	**440**	**301**	**371**
	Std.deviation	*14.28*	*15.48*	*11.49*	*27.71*	*65.76*	*45.71*
	Inner-suburb	**120**	**141**	**208**	**447**	**352**	**375**
	Std.deviation	*18.60*	*26.07*	*71.31*	*46.29*	*87.76*	*61.76*
	Outer-suburb	**70**	**105**	**143**	**392**	**383**	**388**
	Std.deviation	*16.22*	*20.58*	*53.74*	*60.76*	*84.10*	*52.71*
Flat	**Historical-core**	**77**	**181**	**148**	**412**	**585**	**638**
	Std.deviation	*8.92*	*59.27*	*16.49*	*106.80*	*112.03*	*25.09*
	Inner-suburb	**128**	**217**	**167**	**466**	**601**	**647**
	Std.deviation	*22.55*	*46.23*	*22.82*	*65.99*	*51.32*	*18.42*
	Outer-suburb	**63**	**133**	**172**	**451**	**604**	**639**
	Std.deviation	*3.67*	*8.79*	*6.15*	*24.82*	*39.24*	*12.58*
Compound	**Historical-core**	**73**					
	Std.deviation	*2.43*					
	Inner-suburb	**71**					
	Std.deviation	*1.24*					
	Outer-suburb	**50**					
	Std.deviation	*1.11*					

Overall, simulated house rents across the different dwelling types and sizes, except for compound housing shows that house rents tend to be higher at the suburban locations compared to the more central locations of the metropolis. For example, whereas a 5-bedroom detached house attracted a monthly rent amount of GH¢439 at the core locations, a house of similar characteristics cost relatively higher at GH¢442 and GH¢452 in the inner-suburban and outer-suburban locations of the metropolis respectively. In the case of compound housing however, rents on the average is higher in the historical core (GH¢73) than in the inner (GH¢71) and outer-suburban (GH¢50) locations. This could be attributed to the high demand for compound housing, particularly in the inner-city locations among low-income households who constitute the greater percentage of the simulated household population.

Details of house rent formation and evolution resulting from the encoded market transaction rules over time are illustrated for detached dwellings (figure 8.15), semi-detached dwellings (figure 8.16), flats (figure 8.17). For each housing type a graph is produced to illustrate the evolution of house rents for number of bed-rooms ranging from 1 to 5 within the three urban-zones. Overall, the house rent evolution for the different dwelling types reflect continuous adjustments depicted by a rise and fall in rent amounts over time. The upward and downward trends are also the cumulative effect of the encoded market transaction rules where the outcome of competitive transactions for any property on the rental market affect neighbouring properties of similar characteristics. Whereas the falls for instance, is likely the result of households initially submitting bids five percent lower than their willingness to pay, the rise in the subsequent iteration could in part be the result of the learning mechanism where household that are not successful in the previous iteration modify their behaviour by increasing their bids in order to increase their chances of obtaining a property. Thus, upward adjustment of individual bids cumulatively, results in the overall upwards trends in house rents.

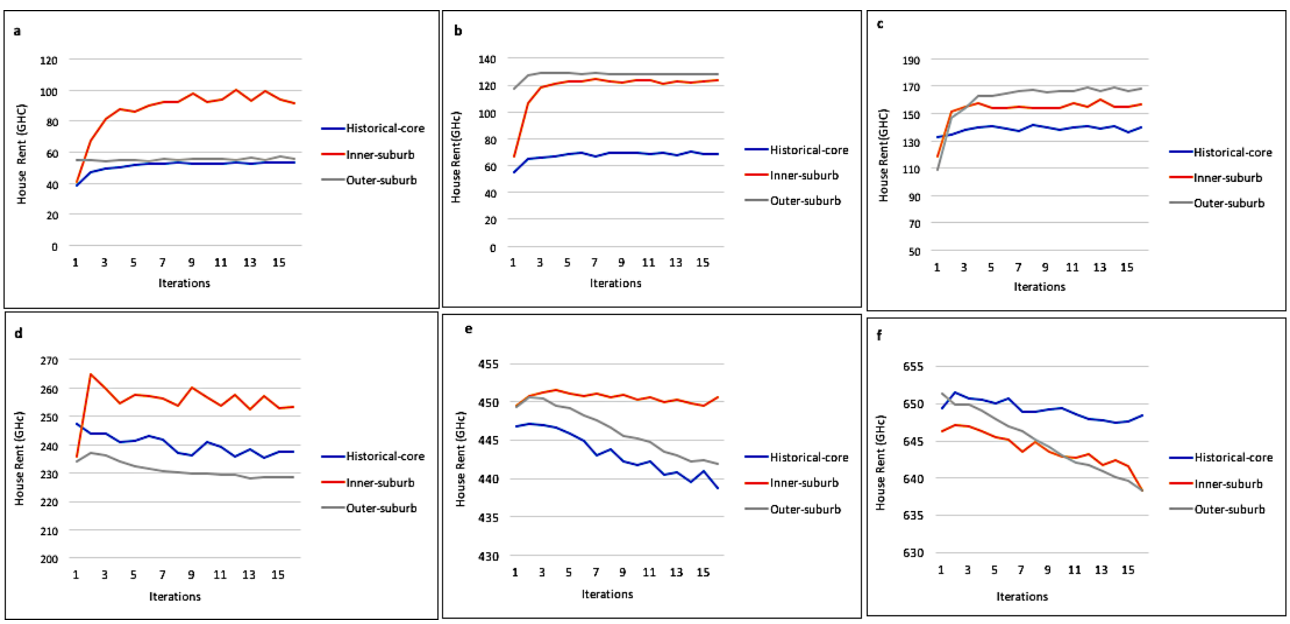


Figure 8.15: Average house rent evolution for detached (a) 1-bedroom (b) 2-bedroom (c) 3-bedroom (d) 4-bedroom (e) 5-bedroom and (f) >5-bedroom house in the historical-core, inner and outer-suburban locations.

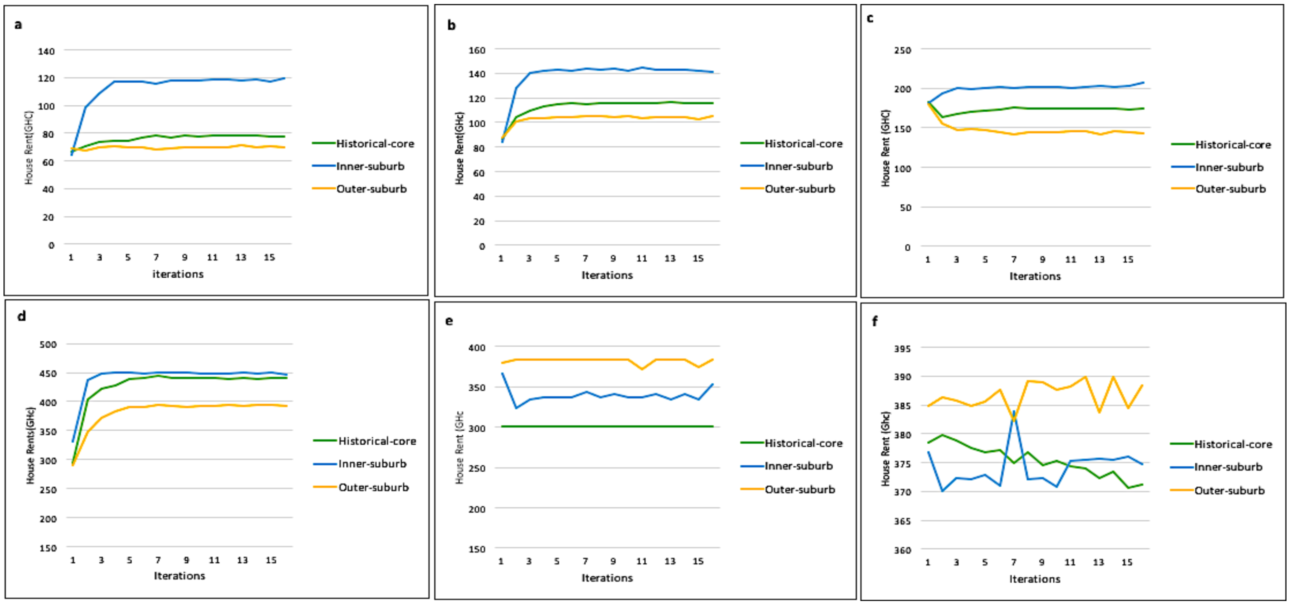


Figure 8.16: Average house rent evolution for semi-detached (a) 1-bedroom (b) 2-bedroom (c) 3-bedroom (d) 4-bedroom (e) 5-bedroom and (f) >5-bedroom house in the historical-core, inner and outer-suburban locations.

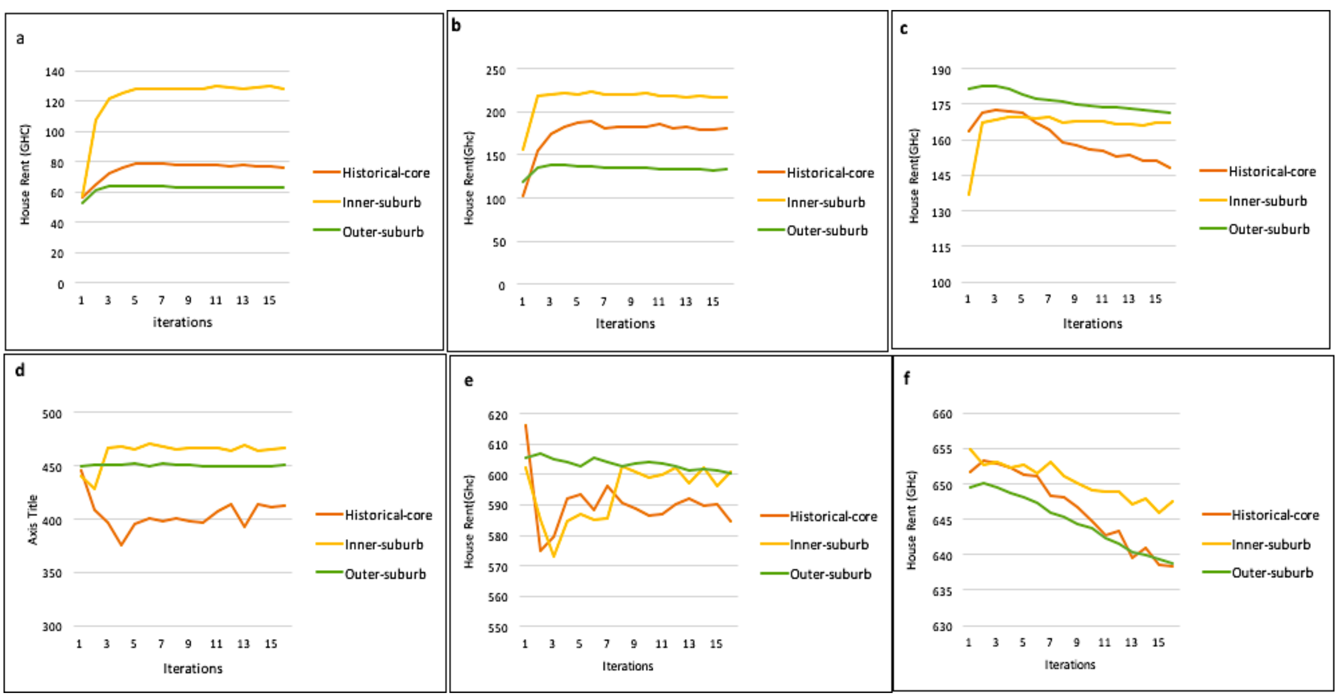


Figure 8.17: Average house rent evolution for flats (a) 1-bedroom (b) 2-bedroom (c) 3-bedroom (d) 4-bedroom (e) 5-bedroom and (f) >5-bedroom house in the historical-core, inner and outer-suburban locations.

A notable trend where rent amount is high initially, but falls over times is also observed in figures 8.15(e), 8.16(f), 8.17(f). These properties are noticeably large size properties of 4 bedrooms and above which also had some of the highest rent levels at the initialization of the model. The observed systematic fall in the rent of these properties could be attributed to the internal price adjustment and learning mechanisms implemented; by this mechanism, a property that remains unlet after being evaluated by the households several times decreases its ask rent to increase the chance of being let. Thus, to avoid the risk of not letting these properties because they are expensive relative to prospective households' willingness to pay and submitted bids, the initial rents are progressively lowered for these properties as illustrated by the results of the simulation. In particular, the spikes in rent beyond the 1st iteration reflect the point where the outcome of the competitive bidding process, resulting from higher bids, leads to average property rents increasing on the market. The fall, on the other hand, reflect the point where prices have increased beyond the amount households are willing to spend on housing, which leads to rents being lowered to meet market power of households.

Property rent evolution within the compound housing sector is depicted in figure 8.18**.** The results show a general trend where average rent in each of the urban zones rise at the initial stages of the simulations but falls smoothly over time.

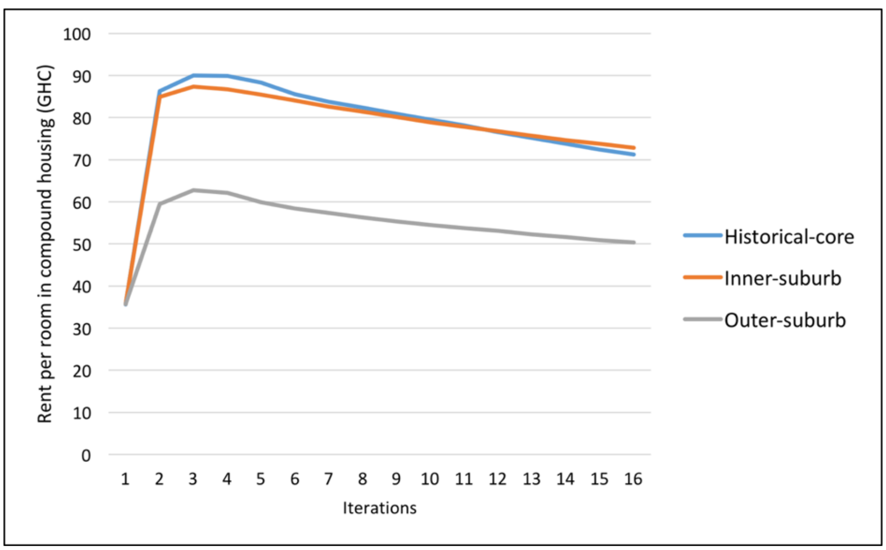


Figure 8.18: Average house rent evolution for a room with a compound house in the historical-core, inner and outer-suburban locations.

The simulated trends reflect the feedback relationship between households submitted bids as determined by their incomes and the prevailing rent levels on the market in the model. Compound housing is the preferred accommodation for low-income households. Given their relatively lower levels of incomes, any initial increases in house rents far beyond their affordability thresholds must be adjusted downwards to match their purchasing power while ensuring that the properties do not remain on the market unlet.

While the formation and evolution of price in the rental market resulted directly from bilateral transactions involving households seeking to rent their accommodation, that of land resulted directly from transaction involving would-be owner-occupier households who, as previously explained, make land acquisition decisions.

Figure 8.19 shows emergent land prices at the end of the simulation based on initial land prices inputted at the model initialization. The resulting land prices are presented in discrete rings of 1km radius from the central locations of the metropolis. The simulated average land price per acre within 1km radius of the metropolitan CBD for example was GH¢222, 272 compared to average simulated land value in areas located within the 10th kilometre buffer of GH¢63,386.

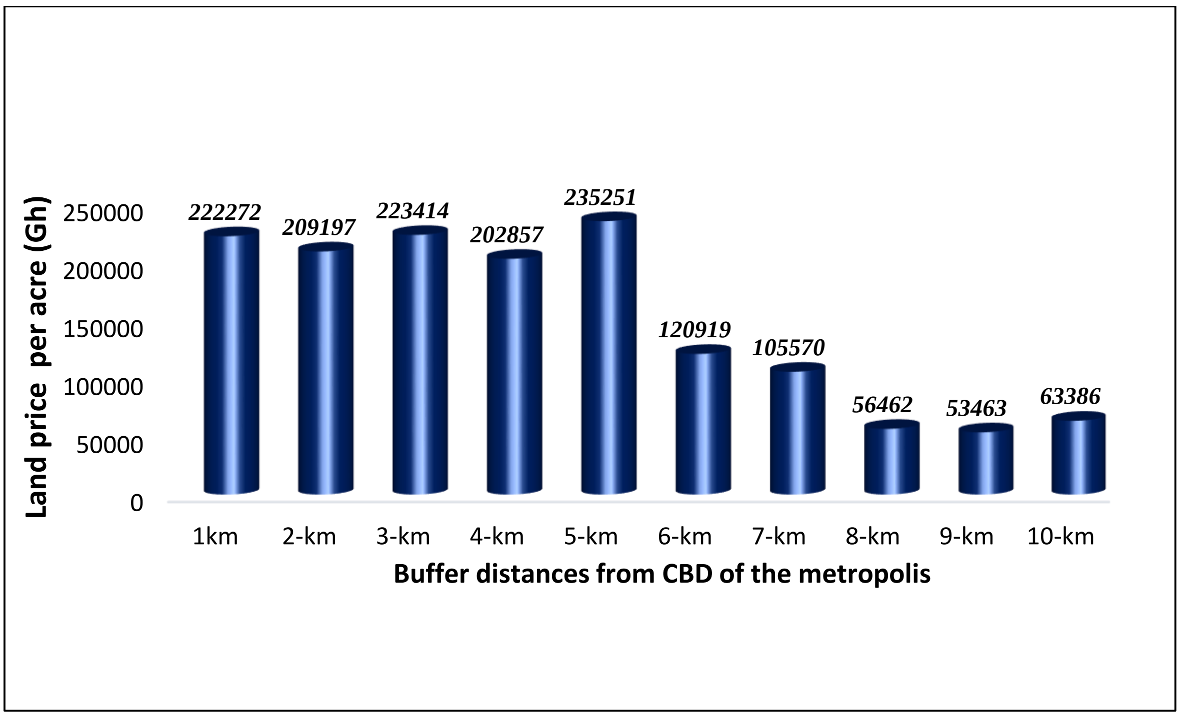


Figure 8.19: Emergent land prices within 1-10km discrete concentric zones from the metropolitan CBD.

Overall, the results show that the highest land prices per acre, at the end of the simulation were recorded in areas located within 5km kilometre radius from the CBD followed by areas with the 6th and 7th kilometre radius and those within the 8th and 10th kilometre radius. In fact, the 1-3km buffer falls within the historical-core of the metropolis, within which the metropolitan CBD is located. Also, Areas within the 4-5km buffer, fall within the high cost residential areas immediately surrounding the CBD in the inner-sub-urban zones of the metropolis: indeed, the two industrial enclaves as well as the University (KNUST) all of which accounted for relatively higher input land values at the model initialization are also located within this zone. Finally, areas within 6-10km radius constitute the outlying suburban locations where land prices are expected to be relatively lower as the simulated results show.

The evolution of the land price over time within each of the 10 discrete concentric zones is shown in figure 8.20. The results show a general increase in land prices over the model simulation period. For example, the average land price within the 1km radius increased from GH¢86,375 on 1st iteration to GH¢222, 272: the average rate of increase across the simulation is 6.50%.

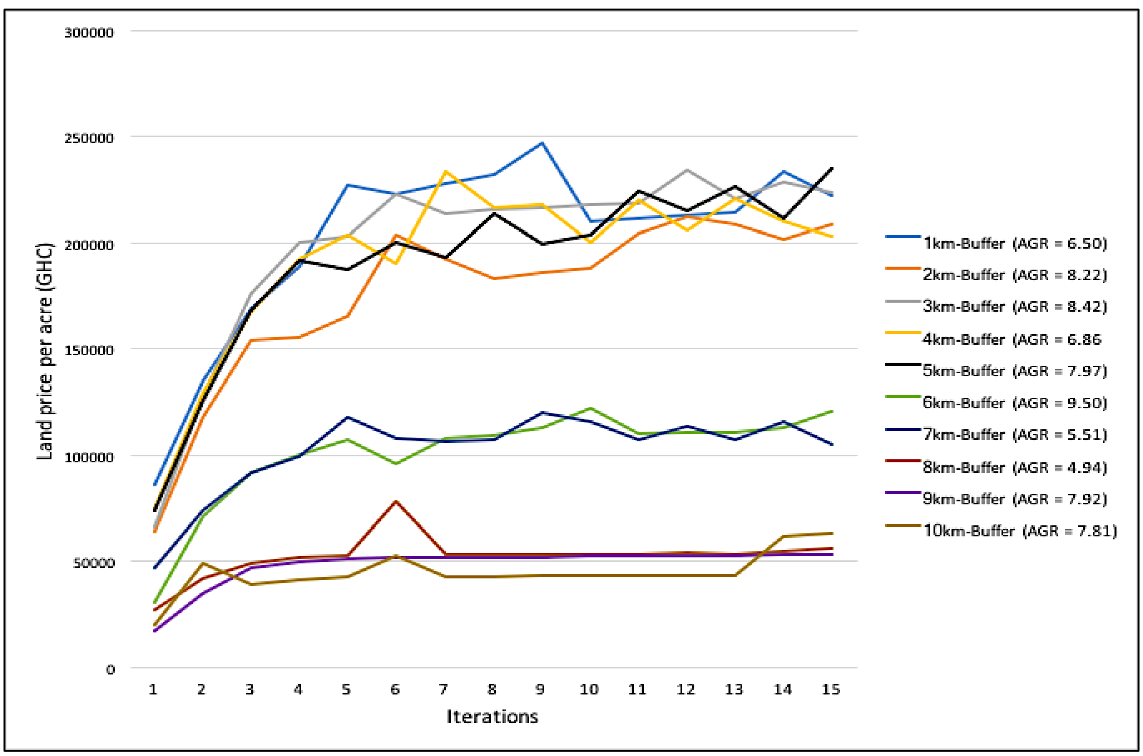


Figure 8.20: The evolution of land price within 1-10km discrete concentric zones from the metropolitan CBD.
Note: AGR refers to the average rate of growth in land prices within each of the zones

Notwithstanding the overall trend of land price increase observed through the simulation, a general trend of surge and fall in land prices is also noticeable in figure 8.20. This upward and downward trend reflect the cumulative effect of the outcome of land transactions with respect to any individual parcel, resulting from the land market competition among owner-occupier households, on the neighbouring parcels of land. Thus, although land prices are expected to increase over time, the mechanism of neighbourhood effect or externalities encoded in the model by which both lower and higher bids at the level of individual parcels cumulatively affect neighbouring parcels, results in upward and downward trends observed globally in the metropolis.

8.3.5 Simulated Job locations and home-work mobility patterns

Besides households making residential location decisions, individual adult members within the households made job location decisions. This section presents model simulation results with respect to job location distributions and the resultant home-work mobility patterns of the individual workers. The simulated job location patterns are presented first, followed by a discussion of the home-work mobility patterns resulting from the simulated residential and job location combinations of workers.

Emergent job location patterns

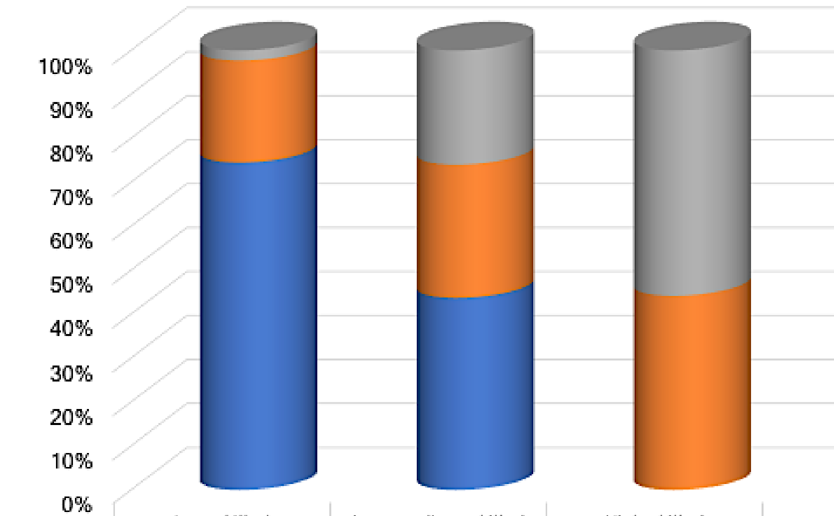
As shown in table 8.13, a total of 63,733 individuals from the households settled in the metropolis attained jobs at the end of the model simulation. Out of this total, 23,995 (38%) individual workers had home-based employment locations meaning that for these workers, their employments are located within the immediate vicinity of their place of residence. The remaining 39739 workers, representing 62% of all workers, had non-home-based job locations. This means that for these individual workers, one of the five major employment zones in the metropolis constituted their job location.

Table 8.13: Simulated job locations of workers in the metropolis.

Iterations	Simulated job locations of workers			Percentages	
	Non-home-based	Home-based	Total	Non-home-based	Home-based
1	16041	9463	25504	63	37
2	24012	14790	38802	62	38
3	28354	17760	46114	61	39
4	31124	19543	50668	61	39
5	33032	20590	53622	62	38
6	34403	21332	55735	62	38
7	35364	21762	57125	62	38
8	36215	22205	58421	62	38
9	36959	22552	59511	62	38
10	37428	22711	60140	62	38
11	38096	23136	61232	62	38
12	38593	23439	62032	62	38
13	38999	23586	62585	62	38
14	39398	23862	63260	62	38
15	39739	23995	63733	62	38

Among individual workers choosing home-based employment, the majority (92%) were low skilled while the remaining eight and two percent belonged to intermediate and high-skilled workers respectively. Moreover, consistent with the observed situation in the case study metropolis, a disproportionately larger share of all home-based employment (71%) generated by the simulation was in the historical-core, followed by the inner-suburban-zone (24%) and outer suburban zone (five percent).

Furthermore, the analysis of home-based employment location examined the distribution of home-based workers in the three urban-zones for low-skilled, intermediate-skilled and high-skilled workers. As shown in figure 8.21, about 74% of all low-skilled home-based employment was in the historical-core: the remaining 23% and two percent of low-skilled home-based employment were in inner and outer-suburban zone respectively. Among intermediate skilled home-based workers, 44%, 30% and 26% were in the historical-core, inner-suburb and outer-suburb respectively. High-skilled home-based workers were located within the inner-suburban (44%) and outer-suburban (56%) zones.



	Low-skilled	Intermediate-skilled	High-skilled
Outer-suburb	536	398	302
Inner-suburb	5281	463	239
Historical-core	16930	668	0

Figure 8.21: Distribution of home-based employment among low-skilled, intermediate-skilled and high skilled workers in the three urban zones

Non-home-based workers obtained employment in one of the five main employment zones within the metropolis as shown in figure 8.22. Overall, the distribution of workers' job locations resulting from the model simulation, reflect the observed patterns in the metropolis. Consistent with the prevailing work location patterns, the largest number of individuals' employment were in the metropolitan CBD followed by the two major industrial enclaves.

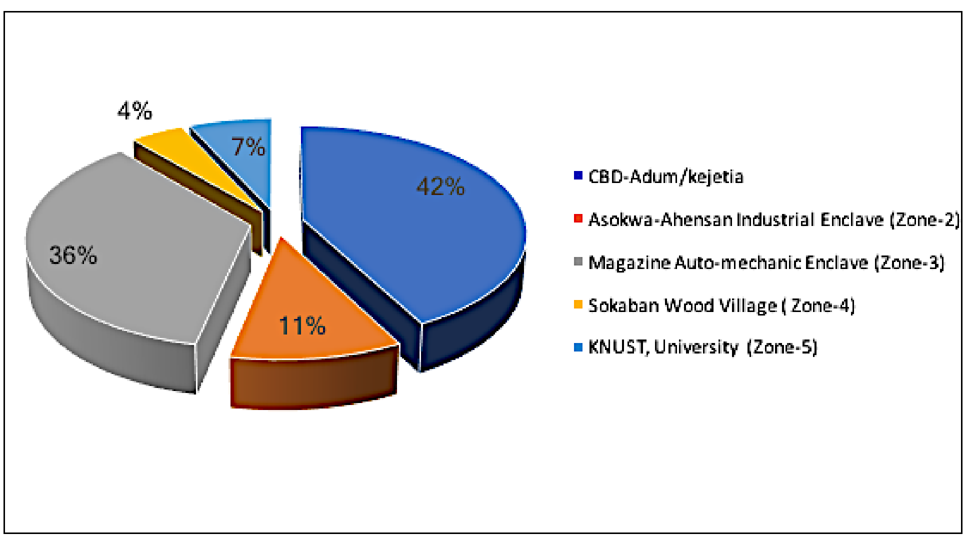


Figure 8.22: Distribution of non-home-based employment among the five major employment zones in the metropolis

As shown in figure 8.22, at the end of the model simulation, the metropolitan CBD accounted for 42% of individual workers' job locations, the largest share of all non-home-based work. The Magazine Auto-mechanic Enclave, designated as employment zone-2 accounted for 36% of non-home-based job locations at the end of the model simulation. Whereas 11% and four percent of all non-home-based workers found employment within the Asokwa-Ahensan Industrial Enclave (employment zone-2) and the Sokoban Wood Village, the wood processing industry designated employment zone 4 respectively, the remaining four percent of attained employments were in the KNUST, the main public university designated as employment zone-5.

Emergent home-work distances and work trip distributions

Besides seeking job opportunities that match their skills, prospective workers in the model sought employment locations with the minimum distance from their home location in order to minimize work travel distance and travel costs. As explained previously, home-based workers having their employments located within or next to their homes avoid having to commute over long distances to work. To non-home-based workers on the other hand, home-work distance separation was important location choice consideration. Table 8.14 provides a summary of the attained distance separation between the work-place and the residence of workers at the end of the simulation. On the average, all non-home-based workers lived some 5,025m (5.03km) from their employment.

Table 8.14: Attained home-work distance among simulated non-home-based workers

Income-Group of workers	**Home-work distance (m)**	
	Mean	std. deviation
Urban-poor	4492.42	2497.84
Low-income	4480.26	2435.52
Lower-middle-income	4937.18	2783.63
Upper-middle-income	5037.28	2763.38
High-income	6100.85	2065.71
Rich	4533.37	1327.29
All	**5025.96**	**157.420**

Overall, the emergent home-work distance separations appear consistent with the encoded work distance minimization objective among households of different income groups. Low-income and urban-poor workers have relatively shorter distance between their places of residence and work-place. For example, urban-poor workers with an average home-work distance of 4492m (4.5km) would experience daily work commutes about 1.6km shorter than high-income workers with an average distance of 6100m (6.1km), and about 0.5km less than the metropolitan average.

In addition to the attained home-work distance separations, the analysis examines work trip distribution among the six macro traffic analysis zones in the metropolis based on simulated home and work locations. The contribution of each of the TAZs[39] to total work trip generation (i.e. trip origin) over time is summarised in table 8.15 while the proportion of trips attracted by each of the TAZs (trip destinations) is presented in table 8.16. A summary O-D matrix showing the contributions of the TAZs as work trip origin and destination zones is presented next in table 8.17.

[39] Detailed description of the TAZs has been presented in chapter five. In view of this, this section refers to them without repeating the description previously provided.

Table 8.15: Simulated work trip origins among traffic analysis zones

Iterations	Simulated Work Trip Origins							Percentages					
	TAZ-301	TAZ-302	TAZ-303	TAZ-304	TAZ-305	TAZ-306	Total	TAZ-301	TAZ-302	TAZ-303	TAZ-304	TAZ-305	TAZ-306
1	278	3112	5011	7656	5609	3489	25154	1	12	20	30	22	14
2	398	4706	7491	11481	8550	5540	38167	1	12	20	30	22	15
3	459	5650	8940	13593	10171	6410	45224	1	12	20	30	22	14
4	508	6321	9594	14729	11518	6964	49634	1	13	19	30	23	14
5	542	6894	10411	15559	11701	7356	52463	1	13	20	30	22	14
6	565	7288	10823	15771	12228	7872	54548	1	13	20	29	22	14
7	588	7604	11015	16596	12393	7721	55917	1	14	20	30	22	14
8	605	7830	10921	17121	12836	7882	57195	1	14	19	30	22	14
9	617	8077	11407	17250	12869	8053	58274	1	14	20	30	22	14
10	627	8195	11142	17509	13313	8102	58889	1	14	19	30	23	14
11	644	8396	11395	17960	13371	8205	59972	1	14	19	30	22	14
12	650	8538	11474	18212	13344	8542	60760	1	14	19	30	22	14
13	662	8661	11624	18350	13644	8365	61306	1	14	19	30	22	14
14	666	8765	11551	18551	13651	8767	61951	1	14	19	30	22	14
15	1269	8996	11534	18721	14354	8860	63733	2	14	18	29	23	14

Table 8.16: Simulated work trip destinations among traffic analysis zones

Iterations	Simulated Work Trip Destinations						Percentages				
	TAZ-301 +302	TAZ-303	TAZ-304	TAZ-305	TAZ-306	Total	TAZ-301 +302	TAZ-303	TAZ-304	TAZ-305	TAZ-306
1	6625	4239	8439	4526	1486	25314	26	17	33	18	6
2	10314	5632	12961	7099	2443	38450	27	15	34	18	6
3	12382	6401	15499	8515	2816	45612	27	14	34	19	6
4	13720	6718	16950	9631	3059	50078	27	13	34	19	6
5	14637	7170	17981	9933	3236	52957	28	14	34	19	6
6	15306	7391	18468	10377	3513	55055	28	13	34	19	6
7	15774	7495	19236	10558	3373	56435	28	13	34	19	6
8	16175	7408	19789	10914	3435	57721	28	13	34	19	6
9	16527	7704	20058	11006	3511	58806	28	13	34	19	6
10	16741	7517	20324	11326	3521	59427	28	13	34	19	6
11	17067	7685	20800	11402	3560	60513	28	13	34	19	6
12	17284	7727	21119	11425	3754	61309	28	13	34	19	6
13	17492	7821	21279	11643	3621	61856	28	13	34	19	6
14	17667	7774	21522	11688	3864	62514	28	12	34	19	6
15	18634	7650	21564	12042	3843	63733	29	12	34	19	6

Table 8.17: Simulated work origin-destination distributions between TAZs

Origin TAZs	**Destination TAZs**					**Total trip Origins**
	TAZ-301 +302	TAZ -303	TAZ -304	TAZ -305	TAZ -306	
TAZ-301 +302	5353	812	2804	1232	64	**10265**
TAZ-303	2979	5274	1908	1352	21	**11534**
TAZ-304	4225	740	12425	1283	48	**18721**
TAZ-305	3868	557	2346	7550	32	**14354**
TAZ-306	2209	268	2081	624	3677	**8860**
Total trip Destinations	**18634**	**7650**	**21564**	**12042**	**3843**	**63733**

As the work trip origins distribution show in table 8.17, TAZ-304, located north of the ring road system in the metropolis accounted for the origins of 18,721 work trips, representing 29% of all work-trip origins. Nearly a quarter of all workers would have TAZ-305 as work trip origin zone since their homes are in this zone. This amounts to a total of 14,354 work trips starting from this zone. Whereas TAZ-303 and TAZ-306 accounted for 18% (i.e. 11,534 trips) and 14% (i.e. 8,860 trips) of work trip origins respectively, TAZs 301 and 302 generated two percent and 14% of all work trips respectively; together, a total of 10,265 work trips originate from these TAZs. The trip origin patterns reflect the dominant land use function of the TAZs. The fact that the commercial and service functions of the metropolis are in TAZs 301 and 302 imply that relatively fewer individuals have their homes located in this zone, hence the relatively lower contributions to total work trip origins. The remaining TAZs are dominantly residential areas and therefore contribute significantly to total work trip origins as the simulation results show.

The next pairs of the work trip distribution emerging from the simulation are the trip destinations. As shown in table 8.17, TAZ 301 and 302, overlapping roughly with the boundaries of the metropolitan CBD attracted 18, 634 trips representing 29% of all work-trips in the metropolis. The highest work trip destination TAZ, is TAZ-304 which attracts 34% of all work trips, the equivalent of 21,564 out of the 63,733 work trips generated daily in the model. The Magazine Auto Mechanic Enclave, one of the major employment zones is in TAZ-304. Furthermore, TAZ-305 within which the Asokwa-Ahensan Industrial area is located attracted 12042 trips, representing 19% of all work trips, and the third highest in the metropolis. Expectedly, TAZs 303 and 306 with dominantly residential uses attracted 12% (i.e. 7, 650 trips) and 6 percent (i.e. 3843 trips) of all work trips generated in the metropolis at the end of the simulation.

Another dimension of the simulated work trip distributions, which takes each of the TAZs as work trip origin zones and examines the proportion each of the other TAZs receives of its trips is presented in figure 8.23. Similar to the observed patterns derived from the empirical analysis, the simulated results show that except for TAZs 305 and 306, all the other TAZs attract the largest share of the work trips they generate. This means that a significant proportion of work trips begin and end in the same TAZ. For example, TAZ-301+302 together have more than half of the work trips they generate, having their destinations in the same zone with the remaining half distributed among the other zones. Similarly, 46% and 66% of work trips having their origins in TAZ-303 and TAZ-304 also have these zones as their destinations. One of the key reasons, as discovered from the empirical analysis is the large number of home-based work in the historical-core and inner-suburban areas of the metropolis. Thus, as home-based work implies travel distance not often exceeding 100m from the home, the associated work trips do not go beyond their respective origin TAZs.

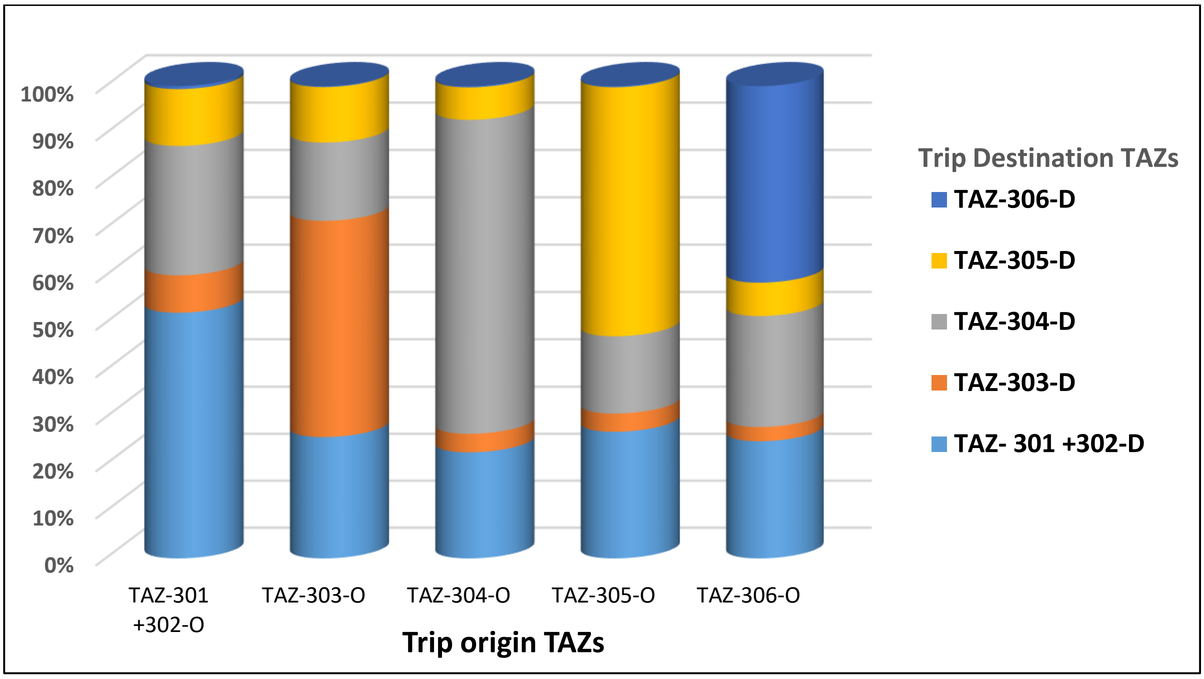


Figure 8.23: Distribution of work trips for each origin TAZ among the destination TAZs

Simulated work travel mode choice

The final aspect of the home-work mobility patterns focuses on work travel mode choice. In the model, travel mode choice to work was simulated probabilistically as function of the income of the workers, the home-work travel distance and associated travel costs. Firstly, two main types of work travel mode choice namely; motorized and non-motorized were simulated. Table 8.18 shows the simulated choice of workers between these two modes.

Table 8.18: Simulated mode choice between motorized and non-motorized transport

Iterations	Simulated mode choice of workers			Percentages	
	Non-motorised	Motorized	Total	Non-motorised	Motorized
1	9463	16041	25504	37	63
2	14793	24017	38809	38	62
3	17765	28359	46125	39	61
4	19550	31130	50680	39	61
5	20597	33037	53635	38	62
6	21340	34408	55749	38	62
7	21770	35369	57139	38	62
8	22214	36221	58435	38	62
9	22560	36965	59525	38	62
10	22720	37434	60154	38	62
11	23145	38101	61246	38	62
12	23448	38598	62046	38	62
13	23595	39004	62599	38	62
14	23871	39403	63274	38	62
15	23986	39747	63733	38	62

The results of choice between motorized and non-motorized work travel mode shows that out of the total of 63733 workers, 39747, representing 62% chose motorized transport while the remaining 23,986 chose non-motorized transport mode (i.e. walking). Indeed, the motorize-non-motorized dichotomy corresponds to the proportion of non-home-based and home-based workers. Most of the non-motorized transport users (i.e. 81%), the simulation results show were workers from low-income households; workers from middle and high income households, accounted for 13% and six percent of non-motorized transport users respectively.

Among workers choosing motorized transport, the model also simulated the probability of choice between two modes namely private transport and public transport. The simulated results show, similar to the observed situation in the metropolis from the empirical analysis that a larger share of workers (i.e. 79%) chose public transport while the remaining 21% of motorized transport users chose private transport. Public transport users further made a probabilistic choice between two sub-categories namely Taxi and Mini-bus/Trotro. Results from the simulation indicate that while 88% of all public transport users chose Mini-bus/Trotro, the remaining 12% opted for Taxi as their daily work travel mode.

Moreover, among workers choosing motorized transport, either public or private as work travel mode, the analysis examines their characteristics in terms of income groupings. The results are presented in figure 8.24. Private car ownership and use as work travel mode at the end of the simulation was found to be higher among workers from high-income (52%) and middle-income

(47%) households; among low-income households simulated ownership and use of the private car as work travel mode was only one percent.

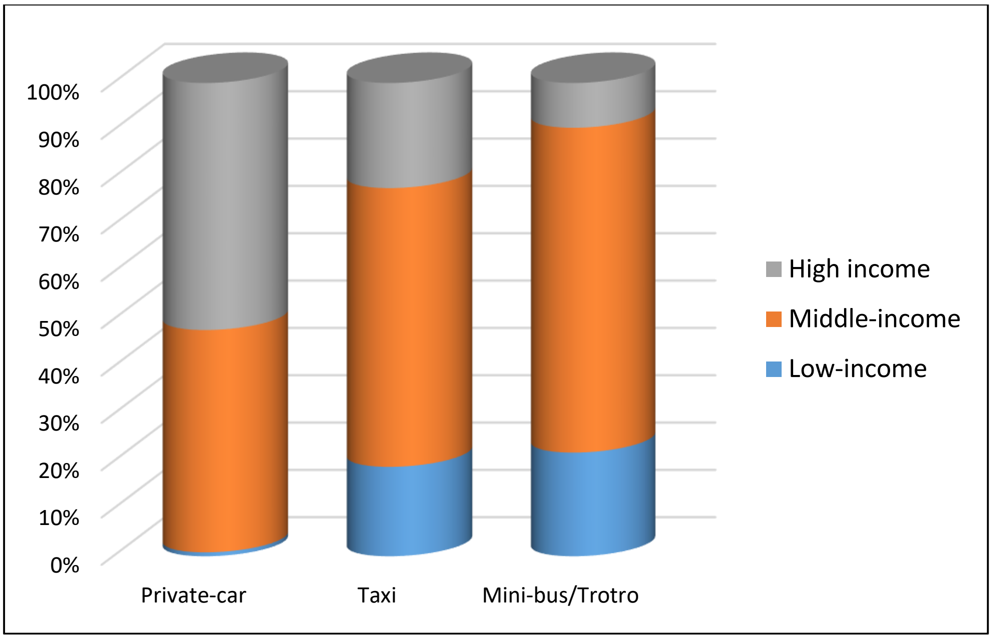


Figure 8.24: Simulated proportions of workers of different income groups and their work travel modes

Furthermore, significantly higher percentage of workers using Taxi as the primary work travel mode were middle-income (59%) and high-income (22%) earners. Consistent with the results of the empirical analysis there were more low-income and middle income households who opted for the Mini-bus/Trotro as work travel model. Together, low-income and middle-income workers constitute 91% of Mini-bus/Trotro users while the remaining nine percent were workers from high income households.

8.4 Chapter Summary

In this chapter, the programmed procedures and commands discussed previously in chapter seven have been used to simulate the residential location patterns of households, the job location patterns of individuals within the household and patterns of mobility associated with the emergent residential-job-location combinations in the Kumasi metropolis.

The model was first calibrated with data from the case study metropolis. As demonstrated, the calibration process involved a series of carefully designed experiments to explore the model parameter space to arrive at best-fit parameter settings for the model. In order to take into account stochastic effects and to reduce uncertainties in outputs of interests generated by the model simulations, an experiment to determine the minimum model run/repetition per iteration using the co-efficient of variation metric calculated for critical model outputs was conducted.

The calibrated model was then used to simulate the urban location and travel choice behaviour of heterogeneous decision makers in the case study metropolis. Overall, the model performed well in replicating simulated patterns which were consistent with calibration criteria specified based on the observational data. The simulated results demonstrated how the encoded feedback relationships between population growth, resulting from the formation of new households based on initial populations, and property market conditions could over time, generate patterns that mimic the urban development process.

Moreover, in terms of residential location patterns at the urban-zone level and the dwelling type, tenancies and amenity proximity levels realized by the simulated households, the simulation generated patterns that could be verified based on the observational data. In addition, results of the simulation showed how the interaction between heterogeneous agents engaged in bilateral transactions and competitive bidding process could account for the formation and evolution of price in the rental housing market as well as in the land market.

The emergent employment location patterns in the model matched closely observed patterns in the target metropolis. Based on the simulated residential and job location combinations, patterns of mobility measured in terms of trip production and attraction patterns between pre-defined Traffic Analysis Zones, home-work distance separation and work mode choice among workers of different socio-economic characteristics were generated that matched the observational data.

In conclusion, this chapter has demonstrated how the integration of theory, survey data and a novel computer-based disaggregate modelling paradigm could reproduce macro-scale urban location and spatial interaction patterns dynamically based on the micro-scale behaviour of purposive households and individuals interacting with each other and their environments. Using a bottom-up approach and holding initial assumptions regarding prevailing urban structural conditions, the approach adopted allows to realistically represent agents as the basis to model and simulate complex choice behaviour processes involving non-linear interactions and feedback mechanisms that shape the overall emergent structure of cities.

The next chapter provides further discussions on the activities undertaken to ensure that the programming of the model was correct (i.e. verification) and that the model adequately represents the metropolitan context as was intended in the conceptual model (i.e. validation).

CHAPTER NINE: MODEL VERIFICATION AND VALIDATION

9.1 Introduction

In chapter six, a conceptual model framework of the co-evolution of urban location and mobility patterns was formulated. The implementation of the conceptual model, involving the translation of decision-making rules, heuristics and internal model dynamics into a computer program was presented in chapter seven, and followed with the model calibration and analyses of simulated results in chapter eight. This chapter focuses on two other important steps in the model development process, which are verification and validation.

The model verification process encompasses all activities conducted in the development of the model with the objective of ensuring that the programming implementation of the conceptual model is correct (Wilensky and Rand, 2007; Xiang et al., 2005). Validation on the other hand, is the process of determining whether the implemented model corresponds to and explains the real-world phenomenon being modelled (North and Macal, 2007). Verification and validation underpin the correctness and utility of the model and are therefore vital to the credibility, validity and replicability of the model.

The transition from programming through calibration to verification and validation of the model is not a linear one. Instead, as demonstrated in chapters seven and eight, at various stages of the programming and the calibration processes of the model, some form of verification and validation took place. Indeed, model calibration and validation are equivalent; calibration often involve validation because the parameter sweeping experiments do aim at producing parameter values that yield optimal relationship between outputs of interests and the observational data (Batty, 2009; Crooks et al., 2008).

This chapter therefore catalogues the most important activities that were conducted to ensure that the model presented in this thesis was implemented correctly while ensuring that the model results matched the residential and job locations patterns as well as mobility characteristics in the case study metropolis for which it was implemented. This chapter is structured into three main sections. Issues bordering on the model verification process are discussed first and followed with a discussion of the activities that went into its validation, highlighting the usefulness as well as limitations in the observational data used. Concluding remarks and reflections on the entire process are presented in the final section of the chapter.

9.2 Verification of METLOMP-SIM

A series of activities were carried out in the implementation of METLOMP-SIM to ensure that the model agents, decision-making rules and environments work as expected in the conceptual model. The key verification activities carried out included the documentation and communication of the model purpose and the overall structure, component testing of sub-model procedures, verifying quantitative representations and calculations constituting the model's initial conditions and software debugging. Details of these processes are outlined in the sections that follow.

9.2.1 Documentation and communication of the model

A well-articulated and carefully documented model, detailing the overall structure, internal dynamics and programming procedures greatly enhances verifiability by both the developer and users of the model (Gilbert, 2008). In view of this, several modelling and programming conventions were followed in the development of METLOMP-SIM. Firstly, to ensure concise and systematic description of the model, the ODD protocol, a standard protocol recommended for uniform description of ABMs was adopted at the model framework specification stage. The protocol, as presented in chapter six outlines the purpose, underlying theory and principles, structure, entities and all the variables in the model. This systematic approach ensured that the overall model was broken down into smaller bits which could easily be translated into computer codes.

Moreover, to make verification of the model easy for users some of which might not be programmers or familiar with the model's programming language, pseudo-codes have been presented where necessary. Agents' behavioural rules and heuristics have all been presented using schematic illustrations which depict the flow, interrelationships and underlying logic of the model's processes. These schematic illustrations could easily be compared with their coded equivalent to determine whether the programming match the rules as described in pseudo-codes. This not only enhances verification but could also assist replication of the model by others either on the platform used or on other modelling platforms.

9.2.2 Programmatic verification and component testing of model sub-components

Programmatic verification involves the use of various procedures to monitor and debug the computer codes on which the model runs with the goal of reducing coding errors. The computer codes of METLOMP-SIM where first checked using Netlogo's inbuilt syntax checker. This inbuilt debugging mechanism checks all lines of code and flags any errors in the code structure for correction. Despite being a useful debugging tool, the syntax checker can only detect syntax errors. Determining whether the codes do exactly what they are intended to do, however, requires additional verification tests.

One way of detecting hidden errors in the programming code that is beyond the detection capabilities of Netlogo's inbuilt syntax checker is Unit testing. Unit or component testing involves writing small tests of codes that checks whether individual units of the computer program are working properly (Gilbert, 2008). This approach was particularly useful in testing if the spatio-environmental sub-component of the model initialized correctly. As explained in chapter seven, the initialization of this sub-component of the model involves importing geospatial data and embedding quantities of the data into Netlogo's world of grid cells using the GIS extension. Due to the un-uniform overlap between layers of the GIS data and that of Netlogo world, it is possible for some patches to register '*Nan'* values (i.e. non-number values). As the name implies, *'Nan'* values are neither numbers nor string that could be interpreted by the software program. Agents are unable to evaluate *'Nan'* value affected cells while mathematical operations cannot be carried out on the affected patches. For these reasons, the model breaks down and returns an error message anytime any of the affected cells are involved in the execution of some procedure. A simple test was written to detect and resolve affected patches.

```
ask patches [
ifelse (value <= 0) or (value >= 0)
 [ set color blue]
 [ set color red]]
```

On importing the GIS data into the model environment, the above lines of code were used to conduct a simple test of searching for all cells/patches with numerical values either less than or equal to zero or greater than or equal to zero and assigning colours to those that meet the condition. Thus, patches/cells that do not meet the condition are left uncoloured; these are the patches with non-numerical values. The values of interest for these patches are set based on the values of their neighbouring patches, in most cases, the mean of value of the neighbouring cells/patches.

In addition to writing a series of unit tests to check whether procedures produce outputs consistent with expectations, visual testing, using colour shading, output monitors, as well as print statements were used to diagnose programming mistakes. For example, amenity proximity indexes computed in GIS and imported into the model were scale-coloured and displayed visually to make sure the right cell values have been imported and embedded into the model environment. The initialization of each of the model sub-components required a series of computations. For example, the socio-demographic sub-model when initialized divides starting households into income groups, and life-cycle stages. For purposes of easy verification and debugging of sub-model procedures, all the model's sub-components were initialized successively independent of each other. While print statements at the end of each initialization procedure were used to gather information about any potential errors, output monitors were used to collect the results of the computation in the form of descriptive statistics such as percentages and measures of central tendency. The model generated computations under each of the sub-components were checked against expected results. Where distributions after the computation were found not to be consistent with expected results, the codes implementing the computation procedures were checked for errors and corrected.

9.2.3 Verifying and modifying behavioural rules and heuristics

Information about the behaviour of agents in METLOMP-SIM were extracted mainly from the empirical data analysis presented in chapters four and five of this thesis. The pieces of information were encoded into behavioural rules and decision-making heuristics used by the agents in their residential-job location and travel choices. In most cases, more than one version of behavioural rules or heuristic for a particular action by the agents was possible. Thus, besides fitting model parameters to the observational data through calibration, it is important to examine how adjustments to rules of behaviour and heuristics affect model outcomes.

Rules of behaviour and heuristics encoded in agent-based models are best verified by carefully monitoring simulation outcomes of individual procedures as well as outcomes from a combination of several procedures, and systematically modifying these rules to arrive at a version that matches reality in the best possible way **(**Cioffi-Revilla, 2013**).** The most important adjustments in heuristics made in the development of METLOMP-SIM are discussed as follows:

One of the key rules driving household agents' decision-making in the model is the spatial search heuristics used to determine their home locations. The property-search-radius, one of the main components of the heuristic, determines how far from a given starting location agents can perceive their choice alternatives and search for discrete location options in the metropolis. The property-search-radius aside constraining options to be evaluated to a reasonable sample, determines the amount of computation performed by the model as the residential location choice procedures are executed by the households. To ensure that neither too many or too few options were evaluated by the households, different search radius limits were tested in which agents were assigned values starting from 1 and increased systematically. For each search radius assigned, a simple test code was written which reported the total number of properties sampled to be evaluated by each household agent. The test results revealed that on the average, a search radius of 1 reported a set of five choice alternatives, and increasing the radius to two, three, five and 10 for example, returned sampled choice sets of 11, 28, 79 and 310 respectively. Given that on the average, households in the case study metropolis indicated that they evaluated six location options in the process of choosing their current places of residence, the search radius of 1, which returns a sample set of five choice alternatives closely matched the reality and was therefore implemented as the search limit.

Implementing the spatial search limit also required one important modification. In many instances, it was found that household agents within the chosen search limit would find locations that met their expected utilities as determined by the amenity proximity values at that location and the preference weights placed on each of the amenities by the households. However, these potential locations may not have the preferred dwelling type which provide the most suitable tenancy arrangement for the household. This affects the success rate of households in obtaining a suitable place of residence, increases the length of residential location search times in the model far beyond realistic time periods while increasing overall simulation time. To avert this, the search heuristic was modified such that household agents would first identify and move to one potential residential location that meets all preferences—amenity proximity preferences, preferred dwelling type and preferred tenancy arrangement within the urban-zone where the search is taking place. From this initial potential location, the household agent samples other properties within the specified search radius, which includes the initial potential suitable location. Implementing this modified version of the search heuristic ensures that each household has at least one option within the choice alternatives that meets it preferences which could become their final home location if they were successful at the competition stage where how much they are willing to pay, based on their income levels, determines winners and losers.

The second major modification to the model occurred after the initial calibration. In the initial model, owing to limitations in the data, tenancy types (i.e. owner-occupier, renting and rent-free) were provided exclusively within a single dwelling on the market. In other words, each dwelling unit at the initialization of the property market sub-model only provided a single tenancy arrangement being it owner-occupier or renting or rent-free. Households, based on their preferences constrained by their income levels were then allocated into these dwellings at the execution of the choice procedures. Using this approach meant that simulated results greatly overestimated the proportion of households who end up in the rental sector while significantly underestimating those in the owner-occupier and rent-free sectors. Also, a disproportionately larger share of households ended in compound housing because the rules underpinning their preferences followed rigidly the observed relationships between socio-economic status and dwelling preferences from the survey data.

To address the above problem, two main adjustments were made to the initial model. Firstly, with the help of past census data and knowledge of the context, some of the dwelling types initially designated either exclusively for rent-free or renting were defined as multiple-tenancy dwellings. This means that these dwellings provided dual tenancies of either renting and owner-occupier or rent-free and owner-occupier, implying that in both cases, there would be a resident-landlord in addition to would-be renters or rent-free tenants. Following this, the scope of residential location search of the household agents was modified to include the dwellings with multiple-tenancies as an additional choice alternative.

Besides initializing dwelling with multiple tenancies, households who end up owning their residence, depending on the size of the property, could decide to sublet part of their property in the rental market or provide rent-free accommodation to low income households who are assumed to be related to the owner as is the case within the metropolitan area represented by the model. Furthermore, preferences for dwelling types and tenancy arrangements were modified from initial rigid specifications where certain households for example would enter the property market with the purpose of seeking only rental housing following findings from empirical analysis. Instead, a more flexible and ordered preference in which household agents follow their preference as revealed through the empirical data initially, but modify the preferences subsequently depending on the prevailing property market situation was implemented. The cumulative effect of these modifications ensured that the simulated outputs of the model matched the observational data in the best possible way.

9.2.4 Sensitivity Analysis

In sensitivity analysis (SA), the influence of varying model inputs on the outputs of interests is explored. Model inputs in sensitivity analysis could be individual parameters or entire groups of parameters, initial values of state variables or different model structures (Schmolke et al., 2010). The approach to SA could be local, in which case the effects of small variations in input factors are examined or global, in which case input factors are varied over broad ranges (Thiele et al., 2014).

Sensitivity testing for the model was carried out in two major ways; firstly, by varying a group of parameters also called parameter sweeping during the model calibration to find best-fit values and secondly by examining the impacts of initial conditions on the simulated outputs of interest. As demonstrated during the model calibration in chapter seven, parameter sweeping experiments were conducted using Netlogo's BehaviorSpace tool. This allowed to systematically alter values of selected parameters and to assess how the changes in parameter values as well as multiple combinations of possible conditions affected the model results. The process also allowed to test the effect of extreme values for each of the parameters on the model results. By comparing the results aggregated over several repetitions to a calibration criteria established based on the observational data, the best set of parameter values were established for the model.

In this section, results of additional experiments conducted with the aim of quantifying the sensitivity of the simulation output to initial conditions of the model is presented. As indicated in chapter seven, the final simulation was initialized with 30,000 households. It is however, important to establish whether given the same best-fit parameter settings and varying the household population size initialized at the beginning of the simulation, the model would provide output of interest consistent with the observational data. Thus, the sensitivity test for this initial condition proceeded on the hypothesis that using the same parameter settings and agents' encoded behaviour, the size of the starting population would influence the nature of interactions among the agents, which in turn, could affect the simulated results as the population is scaled up from a few hundreds to thousands of household agents.

To assess the effect of initial starting population on the simulated outcomes of the model, the population size was varied from an initial size of 1000 and increased systematically in an interval of 5000 to 30,000 at which point the total time of the simulation was deemed reasonable within the scope of this research. This experiment was very useful in determining the time it takes for a simulation of a given size to complete as the total number of households initialized at the beginning of the simulation also determined the amount of computation performed by the model. The model was then run using the best-fit parameters derived through the calibration experiments. The results of the analysis are presented at two levels.

At the first stage of the sensitivity test, the final aggregated results for every value of the initial household population simulated are provided and compared with the calibration criteria as shown in table 9.1. The results show that across the different total number of initial household population simulated, the parameter settings provided stable outputs of interest which were consistent with the established calibration criteria. This means that, overall, varying the size of the initial population yielded simulated results which fell within the expected distributions based on the observational data. It is important to highlight that marginal differences in outputs were identified as expected and this could be due to the stochastic processes in the model itself but not the parameter values settings.

Table 9.1: Model simulation results based on different initializing household population sizes

Sub-component	Outputs of interest	Expected Output range	Initializing household population and simulated results						
			1000	5000	10000	15000	20000	25000	30000
Residential location	Total simulated residential households	-	1507	8407	17026	26683	36039	45147	58217
	Households occupying compound housing	40-54%	51%	51%	48%	44%	48%	43%	43%
	Households occupying detached housing	5-10%	15%	11%	9%	8%	8%	9%	8%
	Households occupying semi-detached housing	9-15 %	7%	6%	12%	8%	9%	8%	9%
	Households occupying flat	30-40 %	27%	32%	31%	39%	35%	39%	40%
	Renting households	40-53 %	23%	21%	21%	20%	22%	22%	40%
	Owner-occupier households	20-30 %	51%	46%	55%	53%	47%	44%	30%
	Rent-free households	20-30 %	26%	33%	24%	27%	31%	34%	30%
	Historical-core zone of residence	25-30 %	62%	55%	51%	45%	43%	42%	28%
	Inner-suburban zone of residence	40-50 %	25%	24%	23%	27%	27%	27%	41%
	Outer-suburban zone of residence	30-40 %	13%	21%	26%	28%	31%	31%	31%
	Average distance of home to shopping/market	100-2084m	1467.74	1495.91	1563.69	1640.74	1614.58	1514.53	1821.41
	Average distance of home to school (households with children)	100-853m	562.69	626.03	566.69	634.58	687.04	606.13	613.14
	Average distance of home to transport terminal	100-3432m	1644.30	1924.06	1693.41	1737.36	1831.46	1964.94	2349.81
	Average distance of home to major roads	100-554m	359.14	471.42	432.86	390.01	433.58	446.52	448.89
Job location	Total simulated workers	-	2417	12099	22312	34639	44456	55186	63733
	Total non-home-based employment locations	60-70%	58%	61%	62%	60%	61%	61%	62%
	Total home-based employment locations	30-40%	42%	39%	38%	40%	39%	39%	38%
	CBD employment locations	40-50%	44%	41%	43%	41%	42%	43%	42%
	Zone2 employment locations	7-12%	7%	12%	10%	11%	11%	11%	11%
	Zone3 employment locations	30-40%	40%	37%	36%	36%	36%	34%	36%
	Zone4 employment locations	2-5%	3%	5%	4%	5%	4%	4%	4%
	Zone5 employment locations	5-7%	7%	6%	7%	7%	7%	7%	7%
	Average home-employment location distance (non-home-based jobs)	100-5000m	4579.80	5020.29	4927.21	4874.55	4930.63	5186.30	5025.96

Table 9.1 continued: Model simulation results based on different initializing household population sizes

Sub-model	Outputs of interest	Expected Output range	Initializing household population and simulated results						
			1000	5000	10000	15000	20000	25000	30000
Home-work distribution (TAZs)	Total work-trip origin TAZ-301	2-5%	1%	2%	1%	1%	1%	1%	2%
	Total work-trip origin TAZ-302	10-15%	13%	13%	13%	12%	13%	16%	14%
	Total work-trip origin TAZ-303	15-25%	20%	18%	14%	14%	14%	26%	18%
	Total work-trip origin TAZ-304	25-30%	33%	29%	28%	42%	37%	27%	29%
	Total work-trip origin TAZ-305	20-25%	21%	20%	16%	15%	26%	20%	23%
	Total work-trip origin TAZ-306	11-15%	11%	19%	28%	16%	8%	11%	14%
	Total work-trip destination TAZ-301 + TAZ-302	30-35%	29%	27%	31%	27%	30%	33%	29%
	Total work-trip destination TAZ-303	10-20%	14%	12%	10%	9%	9%	16%	12%
	Total work-trip destination TAZ-304	20-30%	38%	34%	33%	43%	39%	32%	34%
	Total work-trip destination TAZ-305	12-20%	15%	18%	14%	14%	21%	17%	19%
	Total work-trip destination TAZ-306	5-10%	6%	10%	15%	8%	3%	3%	6%
Mode choice	Motorized transport use	60-70%	58%	61%	62%	60%	61%	61%	62%
	Non-motorized transport use (walking)	30-40%	41%	39%	37%	40%	39%	38%	38%
	Private-car ownership and use	15-20%	26%	22%	21%	22%	22%	22%	21%
	Public transport	80-90%	75%	78%	79%	78%	78%	79%	79%
	Public transport (Trotro/mini-bus) use	80-90%	87%	88%	89%	88%	89%	88%	88%
	Public transport (taxi) use	10-15%	13%	12%	11%	12%	12%	12%	12%

Next, a series of linear regression models in which the simulation input (i.e. initial household population) was set as the independent variable and the outputs of interest as dependent variables were fitted to quantify the effect of the former on the latter. Results of the analysis is presented in table 9.2. The results of the regression analysis reinforce estimates summarized previously in table 9.1.

Table: 9.2: Linear regression estimate of the effect of household population initialized on model outputs

Independent Variables	**Unstandardized Coefficients**				
	B	**Std. Error**	**t**	**p-values**	**R-Square**
Total simulated residential households	1.840	.014	132.132	.000	0.99
intercept	*-798.211*	*251.163*	*-3.178*	*.025*	
Households occupying compound housing	.789	.024	32.523	.000	0.98
intercept	*251.322*	*437.654*	*.574*	*.591*	
Households occupying detached housing	.153	.007	23.309	.000	0.99
intercept	*58.414*	*118.126*	*.495*	*.642*	
Households occupying semi-detached housing	.177	.015	11.537	.000	0.98
intercept	*-195.097*	*276.552*	*-.705*	*.512*	
Households occupying flat	.722	.029	25.220	.000	.996
intercept	*-907.058*	*515.998*	*-1.758*	*.139*	
Renting households	.724	.017	42.617	.000	.999
intercept	*874.884*	*306.497*	*2.854*	*.036*	
Owner-occupier households	.504	.014	36.742	.000	.998
intercept	*-487.793*	*247.233*	*-1.973*	*.106*	
Rent-free households	.612	.020	30.420	.000	.997
intercept	-1171.996	363.032	-3.228	.023	
Historical-core zone of residence	.404	.008	51.662	.000	.999
intercept	*-286.408*	*140.913*	*-2.033*	*.098*	
Inner-suburban zone of residence	.764	.048	15.870	.000	.990
intercept	*921.224*	*868.485*	*1.061*	*.337*	
Outer-suburban zone of residence	.672	.052	13.005	.000	.986
intercept	*-1433.027*	*932.316*	*-1.537*	*.185*	
Average distance of home to shopping/market	.001	.002	.524	.623	.228
intercept	*594.805*	*32.825*	*18.121*	*.000*	
Average distance of home to school	.003	.002	1.391	.223	.528
intercept	*1506.363*	*41.629*	*36.185*	*.000*	
Average distance of home to transport terminal	.017	.007	2.573	.050	.755
intercept	*1619.495*	*119.597*	*13.541*	*.000*	
Average distance of home to major roads	.004	.002	1.956	.108	.658
intercept	382.617	35.538	10.766	.000	
Total simulated workers	*2.147*	*.035*	*61.271*	*.000*	*0.99*
intercept	1133.022	631.902	61.271	.000	
Total non-home-based employment locations	1.324	.013	101.606	.000	0.99
intercept	*544.520*	*234.988*	*2.317*	*.068*	
Total home-based employment locations	.823	.025	33.562	.000	.998
intercept	*588.502*	*442.198*	*1.331*	*.241*	
CBD employment locations	.566	.006	94.890	.000	0.99
intercept	*176.622*	*107.482*	*1.643*	*.161*	
Zone2 employment locations	.152	.003	46.683	.000	.999
intercept	*-33.470*	*58.701*	*-.570*	*.593*	
Zone3 employment locations	.456	.011	39.714	.000	.998
intercept	*401.764*	*206.889*	*1.942*	*.110*	

Table: 9.2 continued: Linear regression estimate of the effect of household population initialized on model outputs

Independent Variables	Unstandardized Coefficients		t	p-values	R-Square
	B	Std. Error			
Zone4 employment locations	.061	.003	22.810	.000	.995
intercept	*-4.292*	*48.349*	*-.089*	*.933*	
Zone5 employment locations	.090	.001	158.179	.000	0.99
intercept	*3.896*	*10.227*	*.381*	*.719*	
Average home-employment location distance	.022	.007	3.368	.020	.833
intercept	*4667.226*	*117.439*	*39.742*	*.000*	
Total work-trip origin TAZ-301	.026	.003	8.266	.000	.965
intercept	*-9.909*	*55.961*	*-.177*	*.866*	
Total work-trip origin TAZ-302	.279	.003	86.454	.000	0.99
intercept	*84.751*	*58.127*	*1.458*	*.205*	
Total work-trip origin TAZ-303	.597	.122	4.887	.005	.909
intercept	*-2164.003*	*2204.056*	*-.982*	*.371*	
Total work-trip origin TAZ-304	.478	.130	3.685	.014	.855
intercept	*2240.820*	*2340.584*	*.957*	*.382*	
Total work-trip origin TAZ-305	652.442	1218.930	.535	.615	.915
intercept	.344	.068	5.088	.004	
Total work-trip origin TAZ-306	.231	.085	2.709	.042	.771
intercept	*1153.224*	*1537.447*	*.750*	*.487*	
Total work-trip destination TAZ-301 + TAZ-302	.600	.005	111.874	.000	0.99
intercept	*175.884*	*96.663*	*1.820*	*.128*	
Total work-trip destination TAZ-303	.402	.079	5.066	.004	.915
intercept	*-1393.099*	*1429.651*	*-.974*	*.375*	
Total work-trip destination TAZ-304	.614	.094	6.549	.001	.946
intercept	*1755.950*	*1690.755*	*1.039*	*.347*	
Total work-trip destination TAZ-305	.327	.042	7.748	.001	.961
intercept	*358.172*	*761.691*	*.470*	*.658*	
Total work-trip destination TAZ-306	.086	.052	1.648	.160	.593
intercept	*746.917*	*943.385*	*.792*	*.464*	
Motorized transport use	.823	.025	33.585	.000	.998
intercept	*590.742*	*442.042*	*1.336*	*.239*	
Non-motorized transport use (walking)	1.324	.013	101.232	.000	0.99
intercept	*546.605*	*235.870*	*2.317*	*.068*	
Private-car ownership and use	.277	.007	37.909		.998
intercept	*211.270*	*131.910*	*1.602*	*.170*	
Public transport	1.047	.009	119.359	.000	0.99
intercept	*335.334*	*158.152*	*2.120*	*.087*	
Public transport (Trotro/mini-bus) use	.926	.007	124.523	.000	0.99
intercept	*275.542*	*134.158*	*2.054*	*.095*	
Public transport (taxi) use	.120	.003	40.271	.000	.998
intercept	*59.859*	*53.921*	*1.110*	*.317*	

Overall, varying the initial input household population size while maintaining the same settings for all other model parameters results in small changes in the model output. For example, increasing the initial household population by one only increases the total number of households occupying compound housing at the end of the simulation by nearly one (i.e. 0. 789). The level of statistical significance and R-squared estimates for each of the regression models are also provided.

9.3 Validating METLOMP-SIM

Validation examines the extent to which the model adequately represents the system being modelled (Hick, 2001). It seeks to establish whether the right model has been implemented by assessing the goodness of fit of the model to data. In this sense, model validation and calibration are equivalent with the latter often involved in the former (Batty, 2009; Crooks et al., 2008).

Validity, is not a binary event in which models could simply be classified as valid or invalid (Crooks et al., 2008). Instead, the validity of an ABM exist along two axes for which different measures of fit could be obtained (Wilenski and Rand, 2015). The first axis considers the level or scale at which the validity is being examined. Here validation involves both micro and macro scale considerations. Micro-scale validation is concerned with ensuring that the characteristics, behaviours and mechanisms encoded into the agents in the model match-up with their real-world equivalent. At the macro-scale, validation asks whether the aggregate emergent properties of the model correspond to aggregate properties in the real-world. The second axis of validation considers the level of detail. Here, there is face validation which examines whether mechanisms and properties of the model look like mechanisms and properties of the real-world and empirical validation—making sure that the model generates data that can be demonstrated to correspond to similar patterns of data in the real-world.

9.3.1 Preliminary validation of METLOMP-SIM

Ideally, the model should be validated using data separate from the observational data used for the model implementation and calibration. Such data however is currently not available to validate the model, except for the survey data collected from the case study area. Moreover, validating the model with survey data poses several limitations since the sample size of the surveyed households does not match the simulated population. Although the sampling approach as discussed in chapter three is representative, validating a model in which some fifty thousand households were simulated with a sample data of 665 households would constitute a big leap fraught with huge uncertainties. In addition, the limitations of the data imply that it would be unreasonable beyond calibration and sensitivity analysis to perform any statistical analysis of the relationship between simulated outputs and real-world situation as the basis of validity.

In view of the above limitations, a macro-scale face validation is presented based on the observational data. This validation considered 38 key aggregate indicators representing model output of interest against exact equivalent values derived from the observational data. Table 9.3 presents the results of the face validation. Comparing aggregate simulated results of each of the validation indictors to their empirical equivalent from the survey data shows that overall, the model reproduces values which in proportionate terms, match closely the real-world situation. Any deviations either below or above the observational data could be attributed to the limitations already highlighted.

Table 9.3: Preliminary model validation results

Sub-component	Validation Indicator	Simulated outputs	Observational data
Residential location patterns	Households occupying compound housing	43%	49%
	Households occupying detached housing	8%	15%
	Households occupying semi-detached housing	9%	15%
	Households occupying flat	40%	21%
	Renting households	40%	50%
	Owner-occupier households	30%	20%
	Rent-free households	30%	34%
	Historical-core zone of residence	28%	27%
	Inner-suburban zone of residence	41%	43%
	Outer-suburban zone of residence	31%	30%
	Average distance of home to shopping/market	1821.41m	2084m
	Average distance of home to school	613.14m	853m
	Average distance of home to transport terminal	2349.81m	3432m
	Average distance of home to major roads	448.89m	554m
Job location patterns	Total non-home-based employment locations	62%	70%
	Total home-based employment locations	38%	30%
	CBD employment locations	42%	48%
	Zone2 employment locations	11%	3%
	Zone3 employment locations	36%	25%
	Zone4 employment locations	4%	2%
	Zone5 employment locations	7%	7%
	Average home-employment location distance	5025.96m	4500m
Home-work distribution (TAZs)	Total work-trip origin TAZ-301	2%	2%
	Total work-trip origin TAZ-302	14%	30%
	Total work-trip origin TAZ-303	18%	25%
	Total work-trip origin TAZ-304	29%	17%
	Total work-trip origin TAZ-305	23%	15%
	Total work-trip origin TAZ-306	14%	11%
	Total work-trip destination TAZ-301 + TAZ-302	29%	50%
	Total work-trip destination TAZ-303	12%	16%
	Total work-trip destination TAZ-304	34%	10%
	Total work-trip destination TAZ-305	19%	12%
	Total work-trip destination TAZ-306	6%	5%
Mode choice	Motorized transport use	62%	76%
	Non-motorized transport use (walking)	38%	33%
	Private-car ownership and use	21%	21%
	Public transport	79%	79%
	Public transport (Trotro/mini-bus) use	88%	93%
	Public transport (taxi) use	12%	7%

9.3.2 Towards robust empirical validation of METLOMP-SIM

Beyond the macro-scale face validation presented in this section, a robust micro-scale empirical validation of the model would be necessary to further quantify the fit between the simulated outputs and available real-world data. Obtaining current data of sample size larger than the survey data used for the model implementation as well as overcoming software limitations in terms of model execution speed which would also allow one to perform simulations with sample populations equivalent to those in the real-world, would also enhance future validation of the model.

Validating the model empirically, would require disaggregate data on each of the 38 key indicators variables outlined in table 9.3. For example, the validation of residential location patterns in the model, would require large sample data on the distribution of households in each of the three urban-zones in the case study metropolis. With data from a larger sample, observed distributions with respect to the location, type, size and tenancy arrangements of dwellings could be compared to those generated by the model. In addition, validity of the model could be ascertained by comparing the profiles of selected households in the model at sampled location to profiles of households living in their real-world equivalent. This approach would help to establish whether indeed, the preference algorithms and rules formulated in the model allocates households in a manner that matches real-world allocations.

In terms of employment location patterns, validating the simulated patterns empirically would require large sample data on the distribution of home-based employment among the urban-zones as well as the distribution of non-home-based jobs among the employment zones located both within and outside of the metropolis. Finally, validating the home-work mobility patterns at the level of the Traffic Analysis Zones will require current data on origin and destination patterns in the metropolis. In the absence of up-to-date O-D data, the model was verified using observed O-D information obtained from some 1158 individual workers in the households. A larger sample of origin-destination survey data would allow one to establish and quantify the extent to which the work flow generation and attraction capacities of the individual zones in the real-world match those produced by the model.

9.4 Chapter summary

This chapter has catalogued the activities undertaken as part of the implementation of the agent-based model of urban location choice and mobility patterns to ensure that the programming implementation of the conceptual model was correct and that the model can produce outputs of interest that match data from the real-world.

The model verification process focused on the programmatic verification and component testing of model sub-components. Here, important steps followed to monitor and debug the computer codes underlying behaviour in the model to reduce coding errors were discussed. The key programmatic verification procedures adopted include the use of unit or component testing procedures to check whether individual sub-components of the model have been implemented correctly, and visual testing using colour shading output monitors and print statements to diagnose and resolve programming errors. The second aspect of verification discussed in this chapter involved systematically verifying encoded heuristics and condition -action-rules against expected model outputs and modifying the model codes to derive results consisted with expected patterns in the real-world.

A sensitivity analysis experiment conducted with the aim to quantify the effect of model input variation on outputs of interests was also presented. Here, the analysis ascertained the extent to which the initial starting household affected simulated outputs using summary descriptive statistics and a series of linear regression models. Overall, the results shown marginal difference in output of interested simulated as a result of varying the initial starting population of the model simulation. The linear regression estimates also provided the basis to estimate plausible results for sample household population sizes that were not simulated in the current model.

The final section of this chapter addressed important issues regarding the validity of the model. A macro-scale face validation method was undertaken in which the results of a set of 38 validation indictors were compared with their empirical data equivalent. While overall, the verification, sensitivity analysis and preliminary validation of the model, based on the available data, showed that the model has been implemented correctly and generated outputs of interest that match closely the observational data, empirical validation of the model, using separate data-sets would be necessary to claim robustness in validity as the basis of using the model for forecasting.

CHAPTER TEN: DISCUSSION AND CONCLUSION

10.1 Introduction

This chapter presents the overall conclusion of the thesis, highlighting the key findings of the research, the significance of the findings and how they relate to previous research. It also discusses the limitations of the research and proposes areas for further research.

The chapter is structured into four main sections. It begins with reiterating the main objectives and questions pursued in this research. The major findings of the research are then discussed in line with the research objectives and questions, emphasizing the significance of the findings and how they relate with previous research. The penultimate section reflects on the limitations of the overall research. In the last section of this chapter, possible ways in which the current research could be extended in the future are proposed.

10.2 Revisiting the research objectives and questions

This thesis has been inspired by some 60 years of academic research into how the spatial distribution of urban activities, resulting from the location decisions of key actors, co-evolve with patterns of spatial interaction in metropolitan areas. As indicated in chapter one and demonstrated further through the review of previous research in chapter two, previous research in the field has maintained a strong linkage between empirical enquiry into urban location and travel choice on the one hand and the development of simulation models to explore dynamically, the complex processes underpinning the co-evolution of urban spatial structure and mobility patterns on the other hand.

Based on the research gaps and current research directions in both the area of empirical enquiry and simulation model development, this thesis took a dual and mutually reinforcing focus in terms of its objectives and outputs. The first objective of the research sought to:

- examine empirically, the location choice behaviour of households and individuals with respect to their residential and job locations, and the mobility patterns associated with the observed home-work location combinations.

From this objective, four fundamental questions were derived and pursued in this thesis. The key empirical research questions were:

i. What are the factors and underlying processes of residential location choice behaviour of heterogeneous households in urban areas?
ii. What are the factors and underlying processes of job location choice behaviour of individual working members of the heterogeneous urban households?
iii. What are the interdependence between the residential location choice and job location choice of the households? and;
iv. What are the mobility patterns associated with the emergent residential-job location combinations?

The second objective addressed in this thesis focused on the development of a computer-based simulation model of urban location choice and mobility patterns. Specifically, the research sought to:

- develop an integrated geospatial and agent-based model to simulate how the residential and job location choice behaviour of heterogeneous households and individuals co-emerge with mobility patterns in a metropolitan area.

The model development was grounded in the following specific questions:

i. How do the socio-demographic characteristics and preferences of heterogonous households and individuals interact with existing urban structural conditions to influence urban location choice behaviour?
ii. How do bilateral transactions, competitive behaviour and interactions among individual actors in the property market lead to the formation and evolution of property prices?
iii. What are the residential location patterns that emerge from the interaction between households 'and individuals' choice behaviour and existing urban structural conditions?
iv. What are the employment location patterns that emerge from the interaction between the attributes of individual working members of the households and prevailing job market conditions?
v. How does the emergent residential and job location combinations and individual-level attributes of agents interact to shape the home-work mobility patterns?

In the opening chapter of the thesis, it was argued that the empirical analysis of location choice and mobility patterns, reflected in the first objective addressed in this research, was crucial because it would offer a detailed understanding of how heterogeneous households and individuals make their long-term urban location choice decisions and how these decisions, in turn, shaped their short-term choices related to daily patterns of mobility between the home and work locations. It was further argued that the insights derived from the empirical studies, would provide a firm foundation to operationalize the second objective of this research, which sought to develop an empirically grounded, dynamic simulation model of the co-evolution of urban location choice and mobility patterns.

10.3 Discussion of findings from the empirical studies

Using a case study design and selecting the Kumasi Metropolis, a medium-size metropolis of nearly two-million inhabitants in Ghana, West Africa as the case study area, the empirical objective and questions of this research were addressed. From the case study area, a cross-sectional survey was conducted to obtain data on residential location choice, job location choice and daily home-work mobility choices and characteristics from a random sample of households and individuals.

In chapter four, analysis of the residential location preferences of households, the job location choice of adult individuals within the household and the interdependence between the two long-term location choices was presented. It was argued that, the traditional approach where studies of residential location choice have presented aggregate spatial zones such as TAZs, census tracks, and city districts as discrete choice alternatives differentiated using some broad measures such as zonal accessibility to employment and amenities (e.g. Pinjari et al., 2011; de Palma et al., 2007; Bhat and Guo, 2007; Pagliara and Wilson, 2010; Andrew and Meen 2006; Zolfaghari et al. 2012) was no longer acceptable. A major limitation of this approach is that, by focusing on aggregate spatial zones, it fails to capture the heterogeneity in the urban environment at the different spatial scales while ignoring the full range of discrete choice alternatives and their attributes (Habib and Miller, 2009; Lee and Waddell, 2010; Bhat, 2015).

In corollary of the above, the analysis of residential location choice, was taken beyond the aggregate zonal discretization of choice alternatives to recognising that the choice process involved both considerations for location-defining attributes at the meso and macro spatial scales as well as dwelling-defining attributes at the micro-scale. Adopting this multi-scale approach, the interaction between attributes defining location as discrete choice alternatives and the socio-demographic attributes households to determine their location preferences were examined.

Firstly, the analysis showed heterogeneity in preferences among the different income-groups of households in relation to various location-defining attributes at the metropolitan and neighbourhood scales. It became evident that households of lower-incomes prioritized proximity to major roads, transport terminals and markets/major shopping areas in their residential location decisions, and attached greater importance to living closer to extended-family relations than high income and rich households. Moreover, the principal component analysis of residential location choice factors distilled the meso and macro-scale residential location choice considerations across all households into four key factors namely; *'proximity to major infrastructure and amenities'*, *'family ties and social networks'*, *'character of neighbourhood'*, and *'proximity to core activity locations'* (i.e. work-place and school) of members of the household.

Furthermore, by representing the three urban-zones (i.e. historical-core, inner-suburban and outer-suburban zones) as discrete choice alternatives and including the complete range of dwelling units (i.e. compound, detached, semi-detached and flat in multi-storey buildings), tenancy types (i.e. renting, rent-free and owner-occupier) and the socio-demographic characteristics of the household in a multinomial logistic regression model, variations in preferences across households were established. The analysis showed that attributes of the households, including income, life-cycle stage, levels of educational attainment and family size determined residential locations at the urban-zone level and choice between dwelling types and tenancy arrangements. Overall, low-income, lower-middle income and upper-middle income households with heads having either no or lower levels of educational attainments, preferred residence in the historical-core of the metropolis, where they lived either rent-free or as renters predominantly in traditional compound housing. On the contrary, residence in owner-occupied, single-family detached and semi-detached housing in suburban locations of the metropolis was common among households of relatively higher incomes, and among couples with children. The large presence of the rent-free sector, a non-market tenure arrangement, constituted one of the unique findings from the case study context. Consistent with findings from the handful of studies that have explored this peculiar phenomenon in detail (e.g. Acheampong, 2016; Gough and Yankson, 2011; Tipple et al., 1997), the finding that a considerable proportion of households lived rent-free in extended-family-owned houses reinforces the strong influence family-ties and social networks has on residential location choice in the case study metropolis.

In addition to the analysis of residential location preferences, the spatial distribution of individuals' employments and the determinants of job location choice was examined. Across all workers, it was established that, proximity of the work place to the place of residence and job availability and salary levels were the two most important job location choice considerations. The large presence of home-based employment was also particularly unique to the Kumasi metropolis. The analysis showed that home-based work location, the phenomenon where the home also doubled as the work-place or employment activities were undertaken within the immediate vicinity of the home, was more pronounced among low-skilled, lower-income workers living in the historical-core neighbourhoods of the metropolis.

Although evidence suggests that generally, inner-city locations with high concentration of commercial uses are less likely to be chosen for residential location (e.g. Pinjari et al, 2011; Waddell et al., 2007), for lower-income households, the need for home-work proximity in order to reduce commuting cost make central locations the most preferred areas of residence. Indeed, in the context of the Kumasi metropolis, the clear majority of home-based employment is in small-scale retail trade in the informal economy. Thus, residing in the historical-core neighbourhoods where commercial activities are concentrated offered several benefits including greater access to markets, reduced or no home-work distance separation resulting in reduced commuting costs, relatively cheaper rents for those renting rooms in compound houses and the absence of rents for those living in extended-family-owned housing.

Furthermore, the interdependence between the residential and job location choice was examined chapter four. Here, the analysis sought to establish empirically, whether residential location choice followed sequentially job location choice and vice versa as held in studies of location choice that adopt the traditional mono-centric assumption, or whether the choice process was not hierarchical but co-joint. Two key findings emerged from the analysis of data.

Firstly, retrospective accounts of residential-job location decisions revealed that most households in their initial location decisions, appear to consider where they would live first, and based on the residential location outcomes, decide where to work. With the home location as the reference point, individuals make job location decisions considering the residential location choice factors alluded to earlier. This finding is corroborated by previous research. For example, Waddell et al., (2007) found that contrary to vast body of literature that often considers the workplace location to be exogenous to residential location choice, there is a general bias towards choosing residential locations first followed by job location decisions. Also, as reported by (Boschmann, 2011), residential decisions, particularly among low-income individuals tend to be influenced by mobility options rather than work location considerations. In the context of the Kumasi metropolis, one could argue that the high incidence of home-based employment, particularly in the historical-core neighbourhoods, suggests a co-joint decision-making process where the decision to live and work in these locations are responsive to each other in equal terms. However, a more plausible explanation, supported by the responses of the households interviewed is that of a conditional choice process where the focus initially is on choosing a place of residence based on affordability, given that most households live rent-free in family accommodation or rent rooms in traditional compound houses where market rents are relatively cheaper. Home-based employment therefore arises subsequently because the lower-skill levels of individuals living in these locations make it impossible for them to find formal sector employment which are also non-home-based. Given the opportunities for small scale retail trade offered by the proximity of these locations to the metropolitan CBD, the dwellings subsequently take on the dual role as a place of residence and place of work.

Secondly, the analysis found that nearly half of all households surveyed had not changed their residential locations over the past seven years. One of the main reasons for residential locations remaining stable over several years was related to the nature of the owner-occupier housing situation in the case study metropolis. Thus, given that, owner-occupier households purchase land and spend several years to develop their housing on incremental basis, they either do not move at all or are less likely to change their residence. Similarly, for most low-income, rent-free tenants within the family-housing sector, residential relocation is rare given that changing residence would imply having to incur housing costs either in the rental or owner-occupier markets. Thus, residential moves, although not generally frequent, when they occurred, were more likely to involve renters than owner-occupiers and rent-free tenants. Job locations on the contrary, had remained unchanged for close to a decade for most individual workers within the households interviewed. Indeed, previous studies have established that frequency of job relocations is often lower than the frequency of residential location changes because mainly because job openings, unlike residential vacancies are very limited and hard to find (Wadell et al., 2007; Lee and Wadell, 2010). Finally, the empirical research showed that job location changes were not triggered by residential moves neither do job location changes lead to a change in residential locations.

In chapter five, the residential-employment-commuting nexus in the case study metropolis was examined by analysing individual workers' home-work mobility characteristics and the determinants of their mobility choices. It was found that people lived about 4.5kms from their work place, implying a relatively shorter home-work distance separation. In fact, among individuals with home-based employment locations, the home-work distance separation was significantly lower, not exceeding one-tenth of a kilometre. The relatively shorter home-work distance separation also meant that for nearly half of all work trips, the origin and destination TAZs were the same. The distribution of home-work trips reflected the functional distribution of land-use activities as the two most central TAZs, which essentially overlapped with the extent of the metropolitan CBD, attracted half of all work trips generated from various home locations in the metropolis.

Furthermore, workers' socio-economic attributes, urban-zone of residential location, job location type, and distance separation between the place of residence and the CBD and the place of residence and work-place determined work travel mode choice. Firstly, the analysis found that higher income levels, larger family size, increasing distance between the place of residence and the metropolitan CBD, and residence in the suburban locations of the metropolis increased the likelihood of car ownership and private car use as work travel mode. Besides the private car, public transport (i.e. mini-bus/Trotro) and walking were the main modes of transport to work. As the logistic regression analysis comparing choice between motorized and non-motorized (i.e. walking) work travel modes shown, lower-income levels, residence in the historical-core of the metropolis and home-based employment location all influenced walking as work travel mode. However, car ownership, home-work distance separation exceeding 300 meters and suburban home-locations all influenced higher preference for motorized transport use. In terms of travel costs, the research found that while motorized transport mode choice to work in general increased commuting costs, relatively higher out-of-pocket travel costs was associated with using the private car and taxis. Also, expenditure on work travel increased with higher incomes due to the preference for private cars or taxis among households of relatively higher incomes compared to households of lower incomes who either walk or use public transport (i.e. mini-bus).

Previous research in different contexts have reported findings similar to the commuting characteristics observed in the Kumasi metropolis. These studies have demonstrated, similar to the findings of the current study that residence in inner-city locations with higher housing or employment densities is associated with lower levels of auto ownership (Pinjari et al, 2011) and relatively shorter non-motorized commuting (Yang and Ferreira, 2008; Næss and Jensen 2004; Kim et al., 2001; Gim, 2013). Also, based on empirical evidence from, Hangzhou Metropolitan Area in China, Næss (2013), found that suburban residents where twice more likely to use motorized mode of transport (i.e. private car and taxi) than their inner-city counterparts. Similarly, Aditjandra et al., (2013) in their study of commuting characteristics in North East England residential neighbourhoods, also found that residents of inner-city traditional neighbourhoods perceived greater opportunities for public transport use and walking than residents of suburban neighbourhoods.

10.4 Discussion of the simulation model development and simulation results

In chapter seven, the programming of the Metropolitan Location and Mobility Patterns Simulator (METLOMP-SIM) was presented followed by the calibration of the model and analysis of simulation results in chapter eight. METLOMP-SIM was developed as a spatially explicit model of the residential-employment-commuting nexus using a disaggregate ABM and geospatial simulation approach. By adopting an integrated geospatial and ABM simulation approach, which allowed to represent households and individuals as heterogeneous decision-makers within a heterogonous urban spatial context, the model responded to an increasingly important methodological need for disaggregate and spatially explicit simulation approaches in the field of urban modelling highlighted in the discussion of research gaps in chapter two.

As a facsimile model grounded in theory, the rules and heuristics mimicking location and mobility choice behaviours were derived from findings of the detailed empirical studies conducted in the Kumasi metropolis, the case study area. Thus, building on the empirical evidence, the model went beyond the static, snapshot empirical analysis of the residential-job location choice and mobility patterns to offer dynamic representation of the choice process, allowing to replicate and to understand the co-emergence of the urban location and mobility patterns resulting from agents’ interactions with one another and their environments.

Drawing on the empirical evidence regarding the residential-job location interdependence, the model explicitly simulated both the residential and job location choice processes, overcoming the exogenous work-place assumption held by existing models which adopt the traditional mono-centric models of location choice. Moreover, METLOMP-SIM responded to the shortcomings in some of the models proceeding it by explicitly integrating property market dynamics and price formation and evolution generated as a function of bilateral transactions and competition in the property markets. The model represented location-defining attributes at the meso and macro scales and specified the full range of discrete choice alternatives, including broad urban-zones (i.e. historical-core, inner-suburban and outer-suburban zones) land parcels with different price values and dwelling units of different sizes and types (i.e. compound, detached, semi-detached and flat) and tenancy arrangements (renting, rent-free and owner-occupier). The inclusion of these elements contributed to a major improvement in the model's realism given that as demonstrated through the literature review in chapter two, many existing models of residential choice, although adopting a disaggregate simulation method, do not represent the full range of discrete choice alternatives and their dimensions as they exist in the real-world (see e.g. Ettema, 2011; Fontaine, 2010; Hosseinali et al., 2013; Murray-Rust et al., 2013).

Besides addressing some of the limitations of existing models and improving realism, the simulated results showed the ability of the model to reproduce macro-scale urban location and spatial interaction patterns dynamically based on the micro-scale behaviour of purposive households and individuals interacting with each other and their environments. The programmed feedback relationships between population growth, resulting from the formation of new households based on initial populations, and property market conditions generate patterns that matched closely the observed residential location patterns in the case study metropolis. Moreover, agents' competition and bilateral transaction in the property market, determined the formation and evolution of property prices in the model, which were found to be consistent with the observed prices in the land and property markets in the case study metropolis.

Finally, the model could reproduce employment location distributions similar to that of the observed patterns in the Kumasi Metropolis. Based on the established residential-home locations, measures of spatial interaction including work trip origin and destinations at the level of traffic analysis zones, home-work distance separations, and travel mode choice at the level of individual commuters were also derived. The model simulations provided reliable estimates of home-based and non-home-based job locations as well as the spatial distribution of non-home-based employments among the major employment zones located within and outside of the metropolis. The estimated origin-destination patterns for the home-work location pairs among TAZs as well as the choice between different travel mode options for work purposes, yielded results that could be verified and validated with the available data.

10.5 Limitations and recommendations for future research

The studies presented in this thesis have limitations worth noting. Firstly, the analysis of residential and job location choice process was based on retrospective account of the choice process obtained through the household survey. This means that the survey could only rely on the memory of the respondents, some of whom had made the location choice decision several years ago. This constitutes one of the limitations of cross-sectional surveys in general because with this approach, information about the phenomenon of interest can only be obtained after it has already taken place. Despite relying on the respondents' retrospective account of their location choice process, the resulting data proved useful in explaining how and why the survey respondents lived and worked at their current locations. Thus, limitations of the data do not invalidate the findings thereof given that as shown through the analysis, the residential and job locations of most of the respondents remained unchanged for several years. That said, in addition to studying location behaviour based on decisions taken in the past as presented in this thesis, it would be useful to study future residential and job relocation behaviours. In this regard, a longitudinal design in which respondents are first interviewed at their current locations and interviewed again after they have made (re)location decisions could reveal deeper insights about the choice process in general, and specifically, how changes in the residential and job locations respond to each other over time.

The simulation model developed in this thesis could also be extended in a number of ways. The current model represents initial residential and job decisions and the associated mobility patterns. However, households, particularly those in the rental market do change their residential locations over time. Similarly, individuals do change jobs and their job locations. Using insights on the triggers of residential and job mobility, the model could be extended by incorporating residential and job relocation heuristics.

Moreover, the distribution of major employment locations in the current model remains unchanged throughout the simulation. This assumption is not a limitation per se given that in many cities, including the Kumasi metropolis for which the model was implemented, major employment zones, once established remain unchanged. It is however the case that new employment zones do emerge in time while existing zones could lose their importance within the system of activity-centre distributions. One of the ways in which the current model could be extended would be to incorporate the mechanisms that could result in the emergence of new employment zones and possibly, the decline of existing economic activity areas. These mechanisms are complex to model and would require better understanding and representation of the feedback relationship between businesses and firms' location and relocation behaviour, population growth and urban development policy.

Furthermore, having modelled the two long-term choice process (i.e. residential and job locations), it would be interesting in the future to include short-term, non-work related choices such as shopping, recreation, and route choice in the model. This would allow to model the full range of choice decisions over different time horizons and at different locations as the basis to derive an activity-based travel pattern in the metropolis.

In terms of its spatial coverage, the current model focused mainly on the areas within the metropolitan boundary. In future, it would be necessary to include outlying areas which do not fall within the administrative boundaries of the metropolis. Expanding the boundaries of the model to the sub-regional and regional contexts would allow to include functional linkages and interactions between the metropolitan core and the outlying districts particularly in terms of employment location functions. This way, the magnitude of movement between the metropolitan core and the outlying areas could be replicated as the basis for forecasting.

One of the important issues discussed in relation to agent-based models of complex systems is equifinality—the principle that in these models a given end state can be reached by many potential means. Applied in the context of the model presented in this thesis, the principle of equifinality implies the possibility that parameter settings different from those selected to run the final model exist that would generate similar patterns. This is particularly relevant given that categorical indicators were used during the calibration of the model. As was described in chapter eight, these parameter values were arrived at through the full factorial parameter sweeping experiments conducted. The parameter settings considered best fit in the end, were chosen by holistically by evaluating the complete sets of the model's outputs of interest against the observational data within the limitations imposed by the computational demands of the simulation. That notwithstanding, it will be useful for future work to explore the issue of equifinality and to determine possible ranges of parameter values that when combined, will essentially yield similar values of output of interest. This will contribute to minimising uncertainties in the model's simulated results.

Ultimately, the utility of METLOMPSIM will depend on the extent to which it can inform practical policy application in the KMA where it has currently been implemented, and in other metropolis where it would be calibrated for in the future. A number of issues will however need resolving before the current model could be considered a fully operational model. As was highlighted in the model implementation and simulation chapters, one of the major challenges with the current model is computational speed during simulation. Netlogo, despite having several capabilities which makes it suitable for the development of models of this kind also has the downside of being slow due to the underlying java architecture. Indeed, because of the challenges with computation, only about 10% of the total household population in the KMA could be simulated in the current model. Consequently, not much can be known about behaviour of the model and the stability of the simulated outputs of interests as the population is scaled up to several thousands of households. Overcoming the challenges with speed will require either compiling a new model from the scratch or exploring other bespoke ABM platforms. Addressing the computational challenge, which in turn, will enable full scale simulation of the model will be a critical step towards METLOMP-SIM becoming operational.

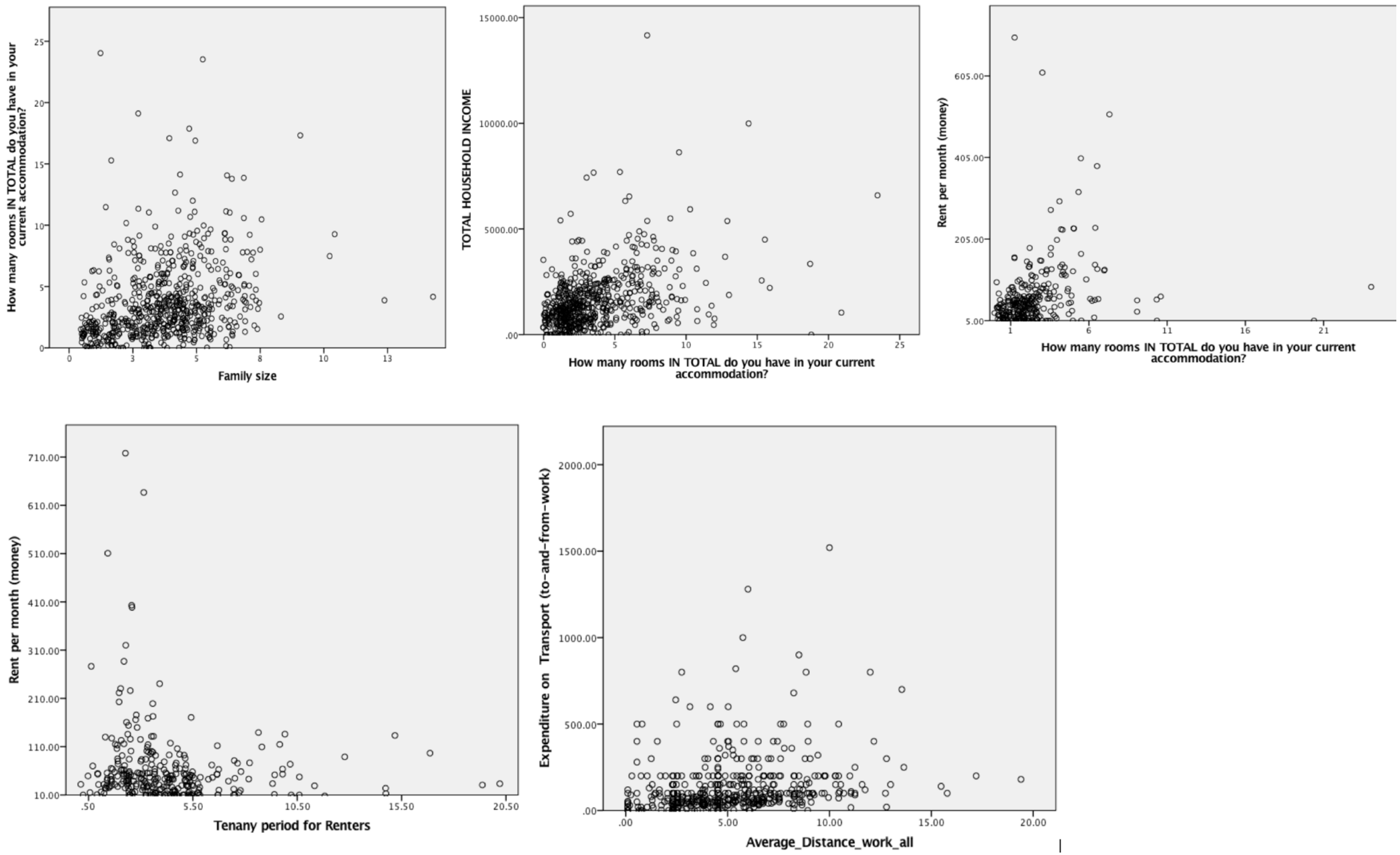
Finally, as demonstrated in chapter nine, the lack of data for validation purposes meant that a preliminary validation of the current model could be conducted using data from the survey data. In the future, a robust micro-scale empirical validation of the model using new set of data on the key validation indicators outlined in chapter eight, would be necessary to further quantify the fit between the simulated outputs and available real-world data. It would be useful to have historical time series data collected in a yearly interval. Such data will be able to capture real-world changes in socio-demographic profiles of households as well as their residential and job location changes against which the outputs of the model, generated on each iteration could be compared. Validating the model with fine scale historical data will also require that other processes that accounts for urban change and changes in location patterns such as gentrification, redevelopment, segregation, firms' location decisions, demographic transitions (i.e. in-migration, out-migration and the formation and dissolution of households) are explicitly captured in the model.

Once the other important urban processes mentioned above are explicitly captured in the model and validated, and a full simulation of the relevant population is achieved, predicting future scenarios of location and mobility patterns, which is not addressed in the current thesis could be undertaken. With respect to residential location choice, preferences of new populations entering the metropolis as a result of the formation of new households and in-migration could be simulated based on the encoded rules of behaviour in the model. Ultimately, the model could be applied to simulating the residential location patterns and travel implications of a master plan for the metropolis or the wider sub-region. In this type of application, the model can take the spatial organization of employment and residential zones proposed by the master plan and simulate the emergent residential-job distributions of the new population expected over the plan period. Another important policy application area that this model could contribute to is the prediction of travel demand. With the work and residential location patterns of the population established and short-term choices with respect to journey times, route choice and daily activity participation explicitly included, total travel demand in the metropolis at different time horizons could be estimated with the model.

10.6 Conclusion

In conclusion, this thesis has contributed to a deeper understanding of how location-defining attributes at multiple scales interact with socio-economic characteristics of heterogeneous households and individuals to determine their residential and job location choices. Positioning the home-work locations as anchors of spatial interaction, this research has also contributed to a deeper understanding of the nature and determinants of spatial flows including travel frequency, trip origins and destinations, transport mode choice, travel times and travel costs. The most significant contribution of this thesis, perhaps, is the development of the Metropolitan Location and Mobility Patterns Simulator (METLOMP-SIM) to simulates the co-emergence of urban spatial structure and mobility patterns as a function of the interaction between individuals' location choice behaviour in the urban property and job markets, and existing urban structural conditions. Facsimile urban models such as the one presented in this thesis derive strength from their firm empirical grounding. The empirical studies provided a solid foundation to integrate the unique attributes of the case study area in the development of the model. Notwithstanding the contextual influences on the model's development, it is possible to extend and apply the current model in different geographical settings through parametrization and calibration with data.

APPENDIX 1: SCATTER PLOT OF VARIABLES IN LINEAR REGRESSION MODELS



BIBLIOGRAPHY

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: *Computational Statistics*, 2(4), 433-459.

Acheampong, R. A. (2016). The family housing sector in urban Ghana: exploring the dynamics of tenure arrangements and the nature of family support networks. *International Development Planning Review*, 38(3), 297-316.

Acheampong, R. A., Agyemang, F. S., & Abdul-Fatawu, M. (2016). Quantifying the spatio-temporal patterns of settlement growth in a metropolitan region of Ghana. *GeoJournal*, 1-18. DOI: 10.1007/s10708-016-9719-x

Acheampong, R. A., & Anokye, P. A. (2013). Understanding households' residential location choice in Kumasi's peri-urban settlements and the implications for sustainable urban growth. *Research on Humanities and Social Sciences, 3(9), 60–70.*

Acheampong, R. A., & Silva, E. (2015). Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, 8(3) 1-28

Adhvaryu, B. (2010). Enhancing urban planning using simplified models: SIMPLAN for Ahmedabad, India. *Progress in Planning*, 73(3), 113-207.

Aditjandra, P. T., Mulley, C., & Nelson, J. D. (2013). The influence of neighbourhood design on travel behaviour: Empirical evidence from North East England. *Transport Policy*, 26, 54-65.

Albeverio, S. (2008). The dynamics of complex urban systems. Heidelberg: Physica-Verlag.

Aljoufie, M., Zuidgeest, M., Brussel, M., van Vliet, J., & van Maarseveen, M. (2013). A cellular automata-based land use and transport interaction model applied to Jeddah, Saudi Arabia. *Landscape and Urban Planning*, 112, 89-99.

Allen, P. M. (2012). Cities and regions as self-organizing systems: models of complexity: Routledge.

Alonso, W. (1964). *Location and land use. Toward a general theory of land rent*. Cambridge, Massachusetts: Harvard University Press.

Amoateng, P., Cobbinah, P. B., & Owusu-Adade, K. (2013). Managing physical development in peri-urban areas of Kumasi, Ghana: A case of Abuakwa. *Journal of Urban and Environmental Engineering (JUEE)*, 7(1).

Amoako, C., & Korboe, D. (2011). Historical development, population growth and present structure of Kumasi. *Future of the Tree: Towards growth and development of Kumasi University Printing Press, KNUST, Kumasi*.

Anas, A. (1982). *Residential Location Markets and Urban Transportation. Economic Theory, Econometrics and Policy Analysis with Discrete Choice Models*. London: Academic Press Incorporated

Anas, A. (1983). *Location and land use. Toward a general theory of land rent*. Cambridge, Massachusetts: Harvard University Press.

Anas, A. (1984). Discrete choice theory and the general equilibrium of employment, housing, and travel networks in a Lowry-type model of the urban economy. *Environment and Planning A*, 16(11), 1489-1502.

Anas, A. (1994). METROSIM: A unified economic model of transportation and land-use. Williamsville, NY: Alex Anas and Associates.

Anderson, C. A., & Titler, M. G. (2014). Development and verification of an agent-based model of opinion leadership. *Implementation Science*, 9(1), 1.

Andersson, J., & Uboe, J. (2012). Some aspects of random utility, extreme value theory and multinomial logit models. Stochastics-*an International Journal of Probability and Stochastic Processes*, 84(2-3), 425-435.

Anderstig, C., & Mattsson, L.-G. (1991). An integrated model of residential and employment location in a metropolitan region. *Papers in Regional Science*, 70(2), 167-184.

Anderstig, C., & Mattsson, L.-G. (1998). Modelling land-use and transport interaction: policy analyses using the IMREL model. In *Network Infrastructure and the Urban Environment*, edited by L. Lundqvist, L. G. Mattsson, and T. Kim. Heidelberg, Berlin: Springer.

Arentze, T., Hofman, F., van Mourik, H., & Timmermans, H. (2000). ALBATROSS: multiagent, rule-based model of activity pattern decisions. *Transportation Research Record: Journal of the Transportation Research Board*, 1706(1), 136-144.

Arentze, T. A., & Timmermans, H. J. P. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7), 613-633.

Arsanjani, J. J., Helbich, M., & Vaz, E. D. (2013). Spatiotemporal simulation of urban growth patterns using agent-based modeling: The case of Tehran. *Cities*, 32, 33-42.

Aschwanden, G. D. P. A., Wullschleger, T., Muller, H., & Schmitt, G. (2012). Agent based evaluation of dynamic city models A combination of human decision processes and an emission model for transportation based on acceleration and instantaneous speed. *Automation in Construction*, 22, 81-89.

Augustijn-Beckers, E.-W., Flacke, J., & Retsios, B. (2011). Simulating informal settlement growth in Dar es Salaam, Tanzania: An agent-based housing model. *Computers, Environment and Urban Systems*, 35(2), 93-103.

Axhausen, K. W., & Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12(4), 323-341.

Babalik-Sutcliffe, E. (2013). Urban Form and Sustainable Transport: Lessons from the Ankara Case. *International Journal of Sustainable Transportation*, 7(5), 416-430.

Barr, S., Fraszczyk, A., & Mulley, C. (2010). Excess travelling-what does it mean? New definition and a case study of excess commuters in Tyne and Wear, UK. European *Transport Research Review*, 2(2), 69-83.

de la Barra, T. D. L. (1989). *Integrated Land Use and Transport Modelling: Decision Chains and Hierarchies*. Cambridge: Cambridge University Press

Department of Urban Roads (2004), *Consultancy Services for Urban Transport Planning and Traffic Management Studies for Kumasi and Tamale*. 2004.Interim Report Vol-2

Barros, J. (2012). Exploring urban dynamics in Latin American cities using an agent-based simulation approach. In *Agent-based models of geographical systems* (pp. 571-589). Netherlands: Springer

Bates, J. (2000). History of demand modelling. *Handbook of transport modelling*, 1, 11-33.

Batty, M. (2007). *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*, Cambridge, MA, and London: The MIT press.

Batty, M. (2013). The new science of cities, Cambridge, MA, and London: MIT Press.

Batty, M. (2016). Classifying urban models. *Environment and Planning B: Planning and Design*, 43(2), 251-256.

Batty, M. (2017). Cities in Disequilibrium. In J. Johnson, A. Nowak, P. Ormerod, B. Rosewell, & Y.-C. Zhang (Eds.), *Non-Equilibrium Social Science and Policy: Introduction and Essays on New and Changing Paradigms in Socio-Economic Thinking* (pp. 81-96). Cham: Springer International Publishing.

Baynes, T., & Heckbert, S. (2009, May). Micro-scale simulation of the macro urban form: Opportunities for exploring urban change and adaptation. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation* (pp. 14-24). Berlin: Springer Heidelberg.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations research*, 30(5), 961-981.

Ben-Akiva, M., & Bowman, J. L. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies*, 35(7), 1131-1153.

Ben-Elia, E., & Shiftan, Y. (2010). Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part a-Policy and Practice*, 44(4), 249-264.

Bernoulli, D. (1968). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5 (1738): 175–192 1730–1731; (Translated into English by L. Sommer, Exposition of a new theory on the measurement of risk, *Econometrica* 22 (1954) 23–36

Bhat, C. R. (2015). A comprehensive dwelling unit choice model accommodating psychological constructs within a search strategy for consideration set formation. *Transportation Research Part B: Methodological*, 79, 161-188.

Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H.-H. (2013). A household-level activity pattern generation model with an application for Southern California. *Transportation*, 40(5), 1063-1086.

Bhat, C. R., & Guo, J. Y. (2007). A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B: Methodological*, 41(5), 506-526.

Bhat, C. R., Guo, J. Y., Srinivasan, S., & Sivakumar, A. (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 1894(1), 57-66.

Bhat, C. R., & Koppelman, F. S. (1999). Activity-based modeling of travel demand. *In Handbook of transportation Science* edited by R. Hall. US: Springer.

Bithell, M., & Brasington, J. (2009). Coupling agent-based models of subsistence farming with individual-based forest models and dynamic models of water distribution. *Environmental Modelling & Software*, 24(2), 173-190.

Blavatskyy, P. R. (2012). Probabilistic choice and stochastic dominance. *Economic Theory*, 50(1), 59-83.

Blavatskyy, P. R., & Pogrebna, G. (2010). Models of Stochastic Choice and Decision Theories: Why Both Are Important for Analyzing Decisions. *Journal of Applied Econometrics*, 25(6), 963-986.

Blumenfeld-Lieberthal, E., & Portugali, J. (2010). Network Cities: A Complexity-Network

Approach to Urban Dynamics and Development. *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, 99, 77-90.

Boarnet, M. G. (2011). A Broader Context for Land Use and Travel Behavior, and a Research Agenda. *Journal of the American Planning Association*, 77(3), 197-213.

Boarnet, M. G., Forsyth, A., Day, K., & Oakes, J. M. (2011). The Street Level Built Environment and Physical Activity and Walking: Results of a Predictive Validity Study for the Irvine Minnesota Inventory. *Environment and Behavior*, 43(6), 735-775.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3) 7280-7287.

Bone, C., Dragicevic, S., & White, R. (2011). Modeling-in-the-middle: bridging the gap between agent-based modeling and multi-objective decision-making for land use change. *International Journal of Geographical Information Science*, 25(5), 717-737.

Boschmann, E. E. (2011). Job access, location decision, and the working poor: A qualitative study in the Columbus, Ohio metropolitan area. *Geoforum,* 42(6), 671-682.

Boulding, K. E. (1956). General systems theory-the skeleton of science. *Management science*, 2(3), 197-208.

Bradley, M., Bowman, J. L., & Griesenbeck, B. (2010). SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. *Journal of Choice Modelling*, 3(1), 5-31.

Bravo, M., Briceno, L., Cominetti, R., Cortes, C. E., & Martinez, F. (2010). An integrated behavioral model of the land-use and transport systems with network congestion and location externalities. *Transportation Research Part B-Methodologic*al, 44(4), 584-596.

Brommelstroet, M. T., & Bertolini, L. (2010). Integrating land use and transport knowledge in strategy-making. *Transportation*, 37(1), 85-104.

Brown, D. G., Riolo, R., Robinson, D. T., North, M., & Rand, W. (2005). Spatial process and data models: Toward integration of agent-based models and GIS. *Journal of Geographical Systems*, 7(1), 25-47.

Brown, D. G., & Robinson, D. T. (2006). Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and Society*, 11(1), 46.

Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., & Sanders, L. (1996). Multiagent systems and the dynamics of a settlement system. *Geographical Analysis*, 28(2), 161-178.

Burguillo-Rial, J. C. (2013). Modeling Complexity: From Cellular Automata to Evolutionary Game Theory and Multiagent Systems with Coalitions. Nostradamus: Modern Methods of Prediction, *Modeling and Analysis of Nonlinear Systems*, 192, 27-28.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C-Emerging Technologies*, 26, 301-313.

Calvetti, D., & Somersalo, E. (2013). Cellular automata, agent-based modeling. *Computation Mathematical Modeling: An Integrated Approach across Scales*, 17, 199-213.

Cao, X. Y., & Fan, Y. L. (2012). Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B-Planning & Design*, 39(3), 459-470.

Cascetta, E. (2009). Random Utility Theory in Transportation Systems Analysis: Models and Applications Boston, MA: Springer US.

Cassel, S. H., Macuchova, Z., Rudholm, N., & Rydell, A. (2013). Willingness to commute long distance among job seekers in Dalarna, Sweden. *Journal of Transport Geography*, 28, 49-55.

Castle, C. J. E., & Crooks, A. T. (2006). Principles and concepts of agent-based modelling for developing geospatial simulations. URL: http://discovery.ucl.ac.uk/3342/1/3342.pdf [access 25th, Feb 2017)

Cervero, R., & Murakami, J. (2010). Effects of built environments on vehicle miles traveled: evidence from 370 US urbanized areas. *Environment and Planning A*, 42(2), 400-418.

Chang, J. S. (2006). Models of the Relationship between Transport and Land-use: A Review. *Transport Reviews*, 26(3), 325-350.

Chen, Y. M., Li, X., Wang, S. J., Liu, X. P., & Ai, B. (2013). Simulating Urban Form and nergy Consumption in the Pearl River Delta Under Different Development Strategies. *Annals of the Association of American Geographers*, 103(6), 1567-1585.

Chowdhury, T. A., Scott, D. M., & Kanaroglou, P. S. (2013). Urban Form and Commuting Efficiency: A Comparative Analysis across Time and Space. *Urban Studies*, 50(1), 191-207.

Cioffi-Revilla, C. (2013). Introduction to computational social science: principles and Applications, US: Springer Science & Business Media.

Clarke, M. I. (1986). Activity modelling-a research tool or a practical planning technique. *Behavioral Research for Transport Policy*, 3-15.

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Economics,* 21(6), 730-741.

Cobbinah, P. B., & Amoako, C. (2012). Urban sprawl and the loss of peri-urban land in Kumasi, Ghana. *International Journal of Social and Human Sciences*, 6(388), e397.

Cooper, C. H. (2017). Using spatial network analysis to model pedal cycle flows, risk and mode choice. *Journal of Transport Geography*, 58, 157-165.

Coppola, P., Ibeas, Á., dell'Olio, L., & Cordera, R. (2013). LUTI model for the metropolitan area of Santander. *Journal of Urban Planning and Development*, 139(3), 153-165.

Crease, P., & Reichenbacher, T. (2013). Linking Time Geography and Activity Theory to Support the Activities of Mobile Information Seekers. *Transactions in GIS*, 17(4), 507-525.

Crecine, J. P. (1964). TOMM. Pittsburgh: Department of City and Regional Planning. RP Technical Bulletin, No. 6.

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems*, 32(6), 417-430.

Curtis, C. (2011). Integrating Land Use with Public Transport: The Use of a Discursive Accessibility Tool to Inform Metropolitan Spatial Planning in Perth. *Transport Reviews*, 31(2), 179-197.

Davidson, W., Donnelly, R., Vovsha, P., Freedman, J., Ruegg, S., Hicks, J., . . . Picado, R. (2007). Synthesis of first practices and operational research approaches in activity-based travel demand modeling. *Transportation Research Part a-Policy and Practice,* 41(5), 464-488.

de Palma, A., Picard, N., & Waddell, P. (2007). Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics,* 62(2), 204-230.

Dong, X., Ben-Akiva, M. E., Bowman, J. L., & Walker, J. L. (2006). Moving from trip-based to activity-based measures of accessibility. *Transportation Research Part A: Policy and Practice*, 40(2), 163-180.

Donnelly, R., & Upton, W. J. (1998). The development of integrated land use-transport models in Oregon. Transportation, land use, and air quality: making the connection. Conference proceedings, May 8, 1998. (pp. 701–710). American Society of Civil Engineer

Du, J., & Wang, Q. (2011). Exploring Reciprocal Influence between Individual Shopping Travel and Urban Form: Agent-Based Modeling Approach. *Journal of Urban Planning and Development-ASCE*, 137(4), 390-401.

Dujardin, S., Pirart, F., Brevers, F., Marique, A. F., & Teller, J. (2012). Home-to-work commuting, urban form and potential energy savings: A local scale approach to regional statistics. *Transportation Research Part a-Policy and Practice*, 46(7), 1054-1065.

Duncan, M. J., Winkler, E., Sugiyama, T., Cerin, E., duToit, L., Leslie, E., & Owen, N. (2010). Relationships of Land Use Mix with Walking for Transport: Do Land Uses and Geographical Scale Matter? *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 87(5), 782-795.

Eboli, L., Forciniti, C., & Mazzulla, G. (2012). Exploring Land Use and Transport Interaction through Structural Equation Modelling. Proceedings of Ewgt 2012 - 15th Meeting of the Euro Working Group on Transportation, 54, 107-116.

Echenique, M., Crowther, D., & Lindsay, W. (1969). A spatial model of urban stock and activity. *Regional Studies*, 3(3), 281-312.

Echenique, M. H., Flowerdew, A. D. J., Hunt, J. D., Mayo, T. R., Skidmore, I. J., & Simmonds, D. C. (1990). The Meplan Models of Bilbao, Leeds and Dortmund. *Transport Reviews*, 10(4), 309-322.

Echenique, M. H., Hargreaves, A. J., Mitchell, G., & Namdeo, A. (2012). Growing Cities Sustainably Does Urban Form Really Matter? *Journal of the American Planning Association*, 78(2), 121-137.

Eluru, N., Pinjari, A. R., Guo, J. Y., Sener, I. N., Srinivasan, S., Copperman, R. B., & Bhat, C. R. (2008). Population updating system structures and models embedded in the comprehensive econometric microsimulator for urban systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2076(1), 171-182.

Epstein, B. (2012). Agent-Based Modeling and the Fallacies of Individualism. *Models, Simulations, and Representations*, 9, 115-144.

Ettema, D. (1996). Activity based travel demand modeling. (Doctor of Philosophy), Technische Universiteit Eindhoven Netherlands. Retrieved from http://alexandria.tue.nl/extra3/proefschrift/PRF13B/9604751.pdf

Ettema, D. (2011). A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems*, 35(1), 1-11.

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, 76(3), 265-294.

Ewing, R., Greenwald, M., Zhang, M., Walters, J., Feldman, M., Cervero, R., . . . Thomas, J. (2011). Traffic Generated by Mixed-Use Developments-Six-Region Study Using Consistent Built Environmental Measures. *Journal of Urban Planning and Development*, 137(3), 248-261.

Chapin, F. S. (1974). Human Activity Patterns in the City: Things People Do in Time and Space. New York: Wiley.

Faraway, J. J. (2016). Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models (Vol. 124): CRC press.

Farooq, B., & Miller, E. J. (2012). Towards integrated land use and transportation: A dynamic disequilibrium based microsimulation framework for built space markets. *Transportation Research Part a-Policy and Practice*, 46(7), 1030-1053.

Feldman, O., Nicoll, J., Simmonds, D., Sinclair, C., & Skinner, A. (2008). Integrated Transportation Land Use Models for Calculations of Wider Economic Benefit in Transport Schemes. *Transportation Research Record* (2076), 161-170.

Feng, J. X., Dijst, M., Prillwitz, J., & Wissink, B. (2013). Travel Time and Distance in International Perspective: A Comparison between Nanjing (China) and the Randstad (The Netherlands). *Urban Studies*, 50(14), 2993-3010.

Field, A. (2014). Discovering statistics using IBM SPSS statistics, London: Sage.

Filatova, T., Parker, D., & Van der Veen, A. (2009). Agent-based urban land markets: agent's pricing behavior, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation*, 12(1), 3.

Filatova, T., Voinov, A., & van der Veen, A. (2011). Land market mechanisms for preservation of space for coastal ecosystems: An agent-based analysis. *Environmental Modelling & Software*, 26(2), 179-190.

Fishburn, P. C. (1989). Non-transitive measurable utility for decision under uncertainty. *Journal of Mathematical Economics*, 18(2), 187-207.

Fontaine, C. M., & Rounsevell, M. D. (2009). An agent-based approach to model future residential pressure on a regional landscape. *Landscape Ecology*, 24(9), 1237.

Ford, A., Barr, S., Dawson, R., Batty, M., & Hall, J. (2010). Integrating a Gis Based Land-

Use Transport Model and Climate Change Impact Assessment Analysis for Sustainable Urban Development of London to 2100. Proceedings of the First International Conference on Sustainable Urbanization (Icsu 2010), 1138-1147.

Forrester, J. W. 1993. System dynamics and the lessons of 35 years. *In A Systems-Based Approach to Policymaking* (pp. 199–240). Springer US.

Fox, M. (1995). Transport planning and the human activity approach. *Journal of Transport Geography,* 3(2), 105-116.

Gainza, X., & Livert, F. (2013). Urban form and the environmental impact of commuting in a segregated city, Santiago de Chile. *Environment and Planning B-Planning & Design*, 40(3), 507-522

Gan, L. P., & Recker, W. (2013). Stochastic Preplanned Household Activity Pattern Problem with Uncertain Activity Participation (SHAPP). *Transportation Science*, 47(3), 439-454.

Gärling, T., Kwan, M.-p., & Golledge, R. G. (1994). Computational-process modelling of household activity scheduling. *Transportation Research Part B: Methodological, 28(5), 355-364.*

Gaube, V., & Remesch, A. (2013). Impact of urban planning on household's residential decisions: An agent-based simulation model for Vienna. *Environmental Modelling & Software,* 45, 92-103.

Geurs, K., Zondag, B., de Jong, G., & de Bok, M. (2010). Accessibility appraisal of land-use/transport policy strategies: More than just adding up travel-time savings. *Transportation Research Part D-Transport and Environment,* 15(7), 382-393.

Geurs, K. T., de Bok, M., & Zondag, B. (2012). Accessibility benefits of integrated land use and public transport policy plans in the Netherlands. *In Accessibility Analysis and Transport Planning: Challenges for Europe and North America*, edited by Karst T. Geurs, Kevin J. Krizek, and Aura Reggiani. Massachusetts, USA: Edward Elgar Publishing, Inc.

Ghana Statistical Service (2013). 2010 Population and Housing Census. National Analytical Report. Url: http://www.statsghana.gov.gh/docfiles/2010phc/National_Analytical_Report.pdf [accessed: April 20, 2017]

Ghana Statistical Service (2000). Population Data Analysis Reports. Socio-economic and Demographic Trends Vol 1 Url: http://catalog.ihsn.org/index.php/catalog/53/download/28487 [accessed: April 20, 2017]

Gilbert, N. (2008). Agent-based models, London, Sage.

Gilbert, N., & Bankes, S. (2002). Platforms and methods for agent-based modeling.

Proceedings of the National Academy of Sciences, 99(suppl 3), 7197-7198.

Gillespie, B. J. (2017). Household Mobility Decisions and Location Choice Household Mobility in America (pp. 89-125), London: Springer.

Gim, T. H. T. (2011). A Comparison of the Effects of Objective and Perceived Land Use on Travel Behavior. *Growth and Change*, 42(4), 571-600.

Gim, T. H. T. (2012). A meta-analysis of the relationship between density and travel behavior. *Transportation*, 39(3), 491-519.

Gim, T. H. T. (2013). The relationships between land use measures and travel behavior: a meta-analytic approach. *Transportation Planning and Technology*, 36(5), 413-434.

Glaeser, E. L., & Resseger, M. G. (2010). The complementarity between cities and skills. *Journal of Regional Science*, 50(1), 221-244.

Goldner, W., Rosenthall, S., & Meredith, J. (1972). rojective land use model PLUM: Theory and application. Berkeley: University of California, Institute of Transportation and Traffic Engineering. Grunfelder, J., and T. S. Nielsen. 2012. Commuting behavior and urban form: A longitudinal study of a polycentric urban region in Denmark. *Geografisk Tidsskrift-Danish Journal of Geography* 112: 2–14.

Gough, K. V., & Yankson, P. (2011). A neglected aspect of the housing market: the caretakers of peri-urban Accra, Ghana. *Urban Studies*, 48(4), 793-810.

Grunfelder, J., & Nielsen, T. S. (2012). Commuting behaviour and urban form: a longitudinal study of a polycentric urban region in Denmark. *Geografisk Tidsskrift-Danish Journal of Geography*, 112(1), 2-14.

Guo, L. A. (2010). Comparison and Analysis of Urban Central Districts' Land Use Mode and Transport System. Icod 2010: Proceedings of the International Conference on Optimization Design, 105-108.

Haase, D., Haase, A., Kabisch, N., Kabisch, S., & Rink, D. (2012). Actors and factors in land-use simulation: The challenge of urban shrinkage. *Environmental Modelling & Software*, 35, 92-103.

Haase, D., Lautenbach, S., & Seppelt, R. (2010). Modeling and simulating residential mobility in a shrinking city using an agent-based approach. *Environmental Modelling & Software*, 25(10), 1225-1240.

Habib, K. M. N., & Kockelman, K. (2008). Modelling choice of residential location and home type: recent movers in Austin Texas. In *87th Annual Meeting of the Transportation Research Board, Washington, DC*.

Habib, M. A., Miller, E. J., & Mans, B. T. (2011). Modeling of Job Mobility and Location

Choice Decisions. *Transportation Research Record* (2255), 69-78.
Hackney, J., & Marchal, F. (2011). A coupled multi-agent microsimulation of social interactions and transportation behavior. *Transportation Research Part a-Policy and Practice*, 45(4), 296-309.
Hancock, D. R., & Algozzine, B. (2015). Doing case study research: A practical guide for beginning researchers: Teachers College Press.
Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, 10(6), 427-444.
Hankey, S., & Marshall, J. D. (2010). Impacts of urban form on future US passenger-vehicle greenhouse gas emissions. *Energy Policy*, 38(9), 4880-4887.
Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2), 73-76.
Hanson, S., & Giuliano, G. (2004). The geography of urban transportation (3rd ed., pp. xii, 419 p.). New York; London: The Guilford Press.
Haynes, P. (2012). Complexity theory and public policy A radical methodology. Public Policy Beyond the Financial Crisis: *An International Comparative Study*, 9, 3-24.
Helminen, V., Rita, H., Ristimaki, M., & Kontio, P. (2012). Commuting to the centre in different urban structures. *Environment and Planning B-Planning & Design*, 39(2), 247-261.
Hensher, D. A., & Ton, T. (2002). TRESIS: A transportation, land use and environmental strategy impact simulator for urban areas. *Transportation*, 29(4), 439-457.
Hess, S., Stathopoulos, A., & Daly, A. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39(3), 565-591.
Hey, J. D., & Panaccione, L. (2011). Dynamic decision making: what do people do? *Journal of Risk and Uncertainty*, 42(2), 85-123.
Hicks, C., & Earl, C. F. (2001). The validation of simulation models. *Assembly Automation*, 21(3), 193.
Hoffmann, C., Abraham, C., White, M. P., Ball, S., & Skippon, S. M. (2017). What cognitive mechanisms predict travel mode choice? A systematic review with meta-analysis. *Transport Reviews*, 1-22.
Horner, M. W., & Schleith, D. (2012). Analyzing temporal changes in land-use transportation relationships: A LEHD-based approach. *Applied Geography*, 35(1-2), 491-498.

Hosmer, D. W., & Lemeshow, S. (2000). Introduction to the logistic regression model. Applied Logistic Regression, Second Edition, 1-30.

Hosseinali, F., Alesheikh, A. A., & Nourian, F. (2013). Agent-based modeling of urban land-use development, case study: Simulating future scenarios of Qazvin city. *Cities*, 31, 105-113.

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, 41(4), 661-689.

Hunt, J. D. (2010). Stated preference examination of factors influencing residential attraction Residential Location Choice (pp. 21-59), London: Springer.

Hunt, J. D., Abraham, J. E., De Silva, D., Zhong, M., Bridges, J., & Mysko, J. (2008). Developing and applying a parcel-level simulation of developer actions in Baltimore. Paper presented at the Transportation Research Board 87th Annual Meeting.

Iacono, M., Levinson, D., & El-Geneidy, A. (2008). Models of transportation and land use change: a guide to the territory. *Journal of Planning Literature*, 22(4), 323-340.

Ibrahim, M. R. (2017). How do people select their residential locations in Egypt? The case of Alexandria. *Cities*, 62, 96-106.

Innocenti, A., Lattarulo, P., & Pazienza, M. G. (2013). Car stickiness: Heuristics and biases in travel choice. *Transport Policy*, 25, 158-168.

Inoa, I. A., Picard, N., & De Palma, A. (2015). Effect of an Accessibility Measure in a Model for Choice of Residential Location, Workplace, and Type of Employment. *Mathematical Population Studies*, 22(1), 4-36.

Janssen, M. A., & Ostrom, E. (2006). Empirically based, agent-based models. *Ecology and Society*, 11(2), 37.

Jia, T., Jiang, B., Carling, K., Bolin, M., & Ban, Y. F. (2012). An empirical study on human mobility and its agent-based modeling. *Journal of Statistical Mechanics-Theory and Experiment.*

Jin, X. B., & White, R. (2012). An agent-based model of the influence of neighbourhood design on daily trip patterns. *Computers Environment and Urban Systems*, 36(5), 398-411.

Jjumba, A., & Dragićević, S. (2012). High resolution urban land-use change modeling: Agent iCity approach. *Applied Spatial Analysis and Policy*, 5(4), 291-315.

Johnson, P. A., & Sieber, R. E. (2011). Negotiating constraints to the adoption of agent-based modeling in tourism planning. *Environment and Planning B-Planning & Design*, 38(2), 307-321.

Johnston, R. A., Shabazian, D. R., & Gao, S. (2003). UPlan: a versatile urban growth model for transportation planning. *Transportation Research Record: Journal of the Transportation Research Board*, 1831(1), 202-209.

Jolliffe, I. (2002). Principal component analysis: Wiley Online Library.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*, 374(2065), 20150202.

Jordan, R., Birkin, M., & Evans, A. (2014). An agent-based model of residential mobility: Assessing the impacts of urban regeneration policy in the EASEL district. *Computers, Environment and Urban Systems*, 48, 49-63.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, 263-291.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, *20*(1), 141-151.

Kaiser, H. F. (1970). A second-generation little jiffy. Psychometrika, 35(4), 401-415.

Kang, C. G., Ma, X. J., Tong, D. Q., & Liu, Y. (2012). Intra-urban human mobility patterns: An urban morphology perspective. *Physica a-Statistical Mechanics and Its Applications*, 391(4), 1702-1717.

Kang, Q. X. (2011). *Dynamic Model of Urban Residential Spatial Structure: House D-S Bid-rent Equilibrium.* 2011 International Conference on Economic, Education and Management (Iceem2011), Vol Ii, 449-452.

Kleinbaum, D. G., & Klein, M. (2010). Maximum likelihood techniques: An overview Logistic regression (pp. 103-127), London: Springer.

Klinger, T., Kenworthy, J. R., & Lanzendorf, M. (2013). Dimensions of urban mobility cultures - a comparison of German cities. *Journal of Transport Geography*, 31, 18-29.

Landis, J., & Zhang, M. (1998). The second generation of the California urban futures model. Part 1: Model logic and theory. *Environment and Planning B*, 25, 657-666.

Landis, J. D. (1994). The California urban futures model: a new generation of metropolitan simulation models. http://www.uctc.net/papers/244.pdf.

Lawe, S., Lobb, J., Sadek, A. W., Huang, S., & Xie, C. (2009). TRANSIMS Implementation in Chittenden County, Vermont Development, Calibration, and Preliminary Sensitivity Analysis. *Transportation Research Record*(2132), 113-121.

Lee, B. H. Y., & Waddell, P. (2010). Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation*, 37(4), 587-601.

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., . . . Parker, D. C. (2015). The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, 18(4), 4.

Lee, K. S., Hobeika, A. G., Zhang, H. B., & Jung, H. J. (2010). Travelers' Response to Value Pricing: Application of Departure Time Choices to TRANSIMS. *Journal of Transportation Engineering-Asce*, 136(9), 811-817. doi:Doi 10.1061/(Asce)Te.1943-5436.0000139

Lemoy, R., Raux, C., & Jensen, P. (2010). An agent-based model of residential patterns and social structure in urban areas. *Cybergeo: European Journal of Geography*.

Leong, W. Y., & Hensher, D. A. (2012). Embedding Decision Heuristics in Discrete Choice Models: A Review. *Transport Reviews*, 32(3), 313-331.

Li, C. N., & Lai, T. Y. (2013). An empirical analysis of links between transport and land use. *Proceedings of the Institution of Civil Engineers-Transport*, 166(2), 105-112.

Li, D. H., & Hua, W. G. (2011). Disaggregate Logit Model of Public Transportation Share Ratio Prediction in Urban City. *Information and Management Engineering*, Pt Vi, 236, 157-165.

Li, X., & Liu, X. (2007). Defining agents' behaviors to simulate complex residential development using multicriteria evaluation. *Journal of Environmental Management*, 85(4), 1063-1075.

Lin, D. Y., Eluru, N., Waller, S., & Bhat, C. (2008). Integration of activity-based modeling and dynamic traffic assignment. *Transportation Research Record: Journal of the Transportation Research Board*, (2076), 52-61.

Lindner, A., Pitombo, C. S., & Cunha, A. L. (2017). Estimating motorized travel mode choice using classifiers: An application for high-dimensional multicollinear data. *Travel Behaviour and Society*, 6, 100-109.

Liu, C., & Shen, Q. (2011). An empirical analysis of the influence of urban form on household travel and energy consumption. *Computers Environment and Urban Systems*, 35(5), 347-357.

Liu, J., Li, R., Zhu, F., & Dong, H. (2011, July). Modeling on Agent-based Artificial Public Transportation. In *Control Conference (CCC), 2011 30th Chinese* (pp. 5596-5600). IEEE.

Loibl, W., & Toetzer, T. (2003). Modeling growth and densification processes in suburban regions—simulation of landscape transition with spatial agents. *Environmental Modelling & Software*, 18(6), 553-563.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, 805-824.

Loomes, G., & Sugden, R. (1987). Some implications of a more general form of regret theory. *Journal of Economic Theory*, 41(2), 270-287.

Lorscheid, I., Heine, B.-O., & Meyer, M. (2012). Opening the 'black box'of simulations: increased transparency and effective communication through the systematic design of xperiments. *Computational and Mathematical Organization Theory*, 18(1), 22-62.

Lowry, I. S. 1964. A Model of Metropolis. Santa Monica, CA: RAND Corporation.

Lu, X., & Liu, Q. (2012). Study on optimization of regional spatial structure based on land use and transport accessibility - A case study of Pearl River Delta region. *Applied Mechanics and Civil Engineering*, 137, 435-439.

Ma, X. S., Liu, X. O., & Lo, H. K. (2010). Maximum Likelihood Estimation of an Integrated Land Use and Transport System. *Transportation and Urban Sustainability*, 621-628.

Ma, X. S., & Lo, H. K. (2012). Modeling transport management and land use over time. *Transportation Research Part B-Methodological*, 46(6), 687-709.

Mackett, R. L. (1990). The systematic application of the LILT model to Dortmund, Leeds and Tokyo. *Transport Reviews*, 10(4), 323-338.

Mackett, R. L. (1991). A Model-Based Analysis of Transport and Land-Use Policies for Tokyo. *Transport Reviews*, 11(1), 1-18.

Mackett, R. L. (1993). Structure of Linkages between Transport and Land-Use. *Transportation Research Part B-Methodological*, 27(3), 189-206.

Magliocca, N., Safirova, E., McConnell, V., & Walls, M. (2011). An economic agent-based model of coupled housing and land markets (CHALMS). *Computers, Environment and Urban Systems*, 35(3), 183-191.

Mahdavi Adeli, A., Adomavicius, G., & Gupta, A. (2017). Toward Understanding the Dynamics of Bidder Behavior in Continuous Combinatorial Auctions: Agent-Based Simulation Approach. Paper presented at the Proceedings of the 50th Hawaii International Conference on System Sciences.

Malayath, M., & Verma, A. (2013). Activity based travel demand models as a tool for evaluating sustainable transportation policies. *Research in Transportation Economics*, 38(1), 45-66.

Manaugh, K., Miranda-Moreno, L. F., & El-Geneidy, A. M. (2010). The effect of neighbourhood characteristics, accessibility, home-work location, and demographics on commuting distances. *Transportation*, 37(4), 627-646.

Maquilon, J. E. C., Gonzalez-Calderon, C. A., & Henao, J. J. P. (2011). The Selection of a Mode of Urban Transportation: Integrating Psychological Variables to Discrete Choice Models. *Dyna-Colombia*, 78(170), 98-105.

Martinez, F. (1996). MUSSA: land use model for Santiago city. *Transportation Research Record: Journal of the Transportation Research Board*, 1552(1), 126-134.

Martinez, F. J. (1992). The Bid Choice Land-Use Model - an Integrated Economic Framework. *Environment and Planning A*, 24(6), 871-885.

Martinez, F. L., & Morales, Y. O. (2012). Agent-Based Simulation Approach to Urban Dynamics *Modeling. Dyna-Colombia*, 79(173), 34-42.

Martínez, L. M., Viegas, J. M., & Silva, E. A. (2007). Zoning decisions in transport planning and their impact on the precision of results. *Transportation Research Record: Journal of the Transportation Research Board*, 1994(1), 58-65.

McConville, M. E., Rodriguez, D. A., Clifton, K., Cho, G., & Fleischhacker, S. (2011). Disaggregate Land Uses and Walking. *American Journal of Preventive Medicine*, 40(1), 25-32.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. *In Frontiers of Econometrics*, edited. by P. Zarembka. New York: Academic Press.

McGregor, D. F., Adam-Bradford, A., Thompson, D. A., & Simon, D. (2011). Resource management and agriculture in the periurban interface of Kumasi, Ghana: Problems and prospects. *Singapore Journal of Tropical Geography*, 32(3), 382-398.

McNally, M. G., & Rindt, C. (2007). The activity approach. *In Handbook of Transport Modeling*, edited by D. A. Hensher, and K. J. Button. Oxford: Pergamon.

Meurs, H., & Haaijer, R. (2001). Spatial structure and mobility. *Transportation Research Part D: Transport and Environment*, 6(6), 429-446.

Miller, E. J., Farooq, B., Chingcuanco, F., & Wang, D. (2011). Historical Validation of Integrated Transport-Land Use Model System. *Transportation Research Record* (2255), 91-99.

Miller, E. J., & Roorda, M. J. (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record: Journal of the Transportation Research Board*, 1831(1), 114-121.

Miller, E. J., & Salvini, P. A. (1998). The integrated land use, transportation, and environment (ILUTE) modeling system: A framework. http://www.civ.utoronto.ca/sect/traeng/ilute/downloads/conference_papers/miller-salvini_trb-98.pdf..

Miller, H. J. (2005). Necessary space-time conditions for human interaction. *Environment and Planning B-Planning & Design*, 32(3), 381-401.

Miller, H. J., & Bridwell, S. A. (2009). A Field-Based Theory for Time Geography. *Annals of the Association of American Geographers*, 99(1), 49-75.

Millward, H., Spinney, J., & Scott, D. (2013). Active-transport walking behavior: destinations, durations, distances. *Journal of Transport Geography*, 28, 101-110.

Mitchell, G., Hargreaves, A., Namdeo, A., & Echenique, M. (2011). Land use, transport, and carbon futures: the impact of spatial form strategies in three UK urban regions. *Environment and Planning A*, 43(9), 2143-2163.

Miyamoto, K., & Udomsri, R. (1996). An analysis system for integrated policy measures regarding land-use, transport and the environment in a metropolis. *Transport, Land-Use and the Environmen*t, 4, 259-280.

Miyamoto, K., & Udomsri, R. (1996). An Analysis System for Integrated Policy Measures Regarding Land-Use, Transport and the Environment in A Metropolis. In Y. Hayashi & J. Roy (Eds.), *Transport, Land-Use and the Environment* (pp. 259-280): Springer US.

Miyamoto, K., V. Vichiensan, V. Sugiki, N. and K. Kitazume. 2007. Applications of RURBAN integrated with a transport model in detailed zone system. http://www.k5.dion.ne.jp/~miyamoto/wctr07_rurban_revised.pdf.

Moeckel, R. (2016). Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. *Journal of Transport and Land Use*, *10*(1)

Moeckel, R., Schürmann, C., & Wegener, M. (2002, 2002). Microsimulation of urban land use.

Mongin, P. (1997). Expected utility theory. *Handbook of economic methodology*, 342-350.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). Introduction to linear regression analysis: John Wiley & Sons.

Moroni, S. (2015). Complexity and the inherent limits of explanation and prediction: Urban codes for self-organising cities. *Planning Theory*, 14(3), 248-267.

Murphy, E. (2012). Urban spatial location advantage: The dual of the transportation problem and its implications for land-use and transport planning. *Transportation Research Part a-Policy and Practice*, 46(1), 91-101.

Murray-Rust, D., Rieser, V., Robinson, D. T., Milicic, V., & Rounsevell, M. (2013). Agent-based modelling of land use dynamics and residential quality of life for future scenarios. *Environmental Modelling & Software*, 46, 75-89.

Naess, P. (2013). Residential location, transport rationales and daily-life travel behaviour: The case of Hangzhou Metropolitan Area, China. *Progress in Planning*, 79, 5-54.

Næss, P. (2015). Built environment, causality and travel. *Transport Reviews*, 35(3), 275-291.

Nam, K., Lim, U., & Kim, B. H. S. (2012). 'Compact' or 'Sprawl' for sustainable urban form? Measuring the effect on travel behavior in Korea. *Annals of Regional Science*, 49(1), 157-173.

Noland, R. B., & Thomas, J. V. (2007). Multivariate analysis of trip-chaining behavior. *Environment and Planning B-Planning & Design*, 34(6), 953-970.

Olaru, D., Smith, B., & Taplin, J. H. E. (2011). Residential location and transit-oriented development in a new rail corridor. *Transportation Research Part a-Policy and Practice*, 45(3), 219-237.

Otter, H. S., van der Veen, A., & de Vriend, H. J. (2001). ABLOoM: Location behaviour, spatial patterns, and agent-based modelling. *Journal of Artificial Societies and Social Simulation*, 4(4).

Paez, A. (2013). Mapping travelers' attitudes: does space matter? *Journal of Transport Geography*, 26, 117-125.

Pagliara, F., Preston, J., & Simmonds, D. (2010). Residential location choice: Models and Applications, London: Springer Science & Business Media.

Park, B. B., & Kwak, J. (2011). Calibration and Validation of TRANSIMS Microsimulator for an Urban Arterial Network. *Ksce Journal of Civil Engineering*, 15(6), 1091-1100.

Parker, D. C., & Filatova, T. (2008). A conceptual design for a bilateral agent-based land market with heterogeneous economic agents. *Computers, Environment and Urban Systems*, 32(6), 454-463.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers*, 93(2), 314-337.

Parry, H. R., & Bithell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling Agent-based models of geographical systems, London: Springer.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). The adaptive decision maker: Cambridge University Press.

Pendyala, R. M., Kitamura, R., Kikuchi, A., Yamamoto, T., & Fujii, S. (2005). Florida activity mobility simulator: overview and preliminary validation results. *Transportation Research Record: Journal of the Transportation Research Board*, 1921(1), 123-130.

Pendyala, R. M., Konduri, K. C., Chiu, Y. C., Hickman, M., Noh, H., Waddell, P., . . .

Gardner, B. (2012). Integrated Land Use-Transport Model System with Dynamic Time-Dependent Activity-Travel Microsimulation. *Transportation Research Record* (2303), 19-27.

Peters, P., Kloppenburg, S., & Wyatt, S. (2010). Co-ordinating Passages: Understanding the Resources Needed for Everyday Mobility. *Mobilities*, 5(3), 349-368.

Pfaffenbichler, P. (2011). Modelling with Systems Dynamics as a Method to Bridge the Gap between Politics, Planning and Science? Lessons Learnt from the Development of the Land Use and Transport Model MARS. *Transport Reviews*, 31(2), 267-289.

Pfaffenbichler, P., Emberger, G., & Shepherd, S. (2010). A system dynamics approach to land use transport interaction modelling: the strategic model MARS and its application. *System Dynamics Review*, 26(3), 262-282.

Pinjari, A. R., & Bhat, C. R. (2011). Activity-based travel demand analysis. *A Handbook of Transport Economics*, 10, 213-248.

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., & Waddell, P. A. (2011). Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6), 933-958.

Putman, S. H. (1991). Integrated Urban Models 2. New Research and Applications of Optimization and Dynamics. London: Pion.

Putman, S. H. (1998). Results from implementation of integrated transportation and land use models in metropolitan regions Network Infrastructure and the Urban Environment (pp. 268-287), London: Springer.

Railsback, S. F., & Grimm, V. (2011). Agent-based and individual-based modeling: a practical introduction: Princeton University Press.

Railsback, S. F., Lytinen, S. L., & Jackson, S. K. (2006). Agent-based simulation platforms: Review and development recommendations. *Simulation*, 82(9), 609-623.

Rand, W., & Rust, R. T. (2011). Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3), 181-193.

Rasouli, S., & Timmermans, H. (2013). Using emulators to approximate predicted performance indicators of complex microsimulation and multiagent models of travel demand. *Transportation Letters*, 5(2), 96-103.

Rasouli, S., & Timmermans, H. (2014a). Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research. *Travel Behaviour and Society*, *1*(3), 79-90.

Rasouli, S., & Timmermans, H. (2014b). Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences*, 18(1), 31-60.

Recker, W. W. (2001). A bridge between travel demand modeling and activity-based travel analysis. Transportation Research Part B-Methodological, 35(5), 481-506.

Recker, W. W., McNally, M. G., & Root, G. S. (1986). A model of complex travel behavior: Part II—an operational model. *Transportation Research Part A: General,* 20(4), 319-330.

Regenwetter, M., & Davis-Stober, C. P. (2012). Behavioral Variability of Choices Versus Structural Inconsistency of Preferences. *Psychological Review*, 119(2), 408-416.

Ricardo, D. (1821). On the principles of political economy, and taxation: John Murray.

Richards, K., Bithell, M., & Bravo, M. (2004). Individuals, emergence and geographies of space and place. Unifying geography: common heritage, shared future, 327.

Robinson, D. T., Brown, D. G., Parker, D. C., Schreinemachers, P., Janssen, M. A., Huigen, M., . . . Irwin, E. (2007). Comparison of empirical methods for building agent-based models in land use science. *Journal of Land Use Science*, 2(1), 31-55.

Robinson, D. T., Murray-Rust, D., Rieser, V., Milicic, V., & Rounsevell, M. (2012). Modelling the impacts of land system dynamics on human well-being: Using an agent-based approach to cope with data limitations in Koper, Slovenia. *Computers Environment and Urban Systems*, 36(2), 164-176.

Ronald, N. (2013). Agent-Based Modelling of Socio-Technical Systems (Agent-Based Social Systems). *JASSS-The Journal of Artificial Societies and Social Simulation*, 16(2).

Roukouni, A., Basbas, S., & Kokkalis, A. (2012). Impacts of a metro station to the land use and transport system: the Thessaloniki Metro case. *Transport Research Arena* 2012, 48, 1155-1163.

Salter, R. M. (2013). Nova: A modern platform for system dynamics, spatial, and agent-based modeling. 2013 International Conference on Computational Science, 18, 1784-1793.

Samet, R. H. (2013). Complexity, the science of cities and long-range futures. *Futures*, 47, 49-58.

Sanders, L., Pumain, D., Mathian, H., Guérin-Pace, F., & Bura, S. (1997). SIMPOP: a multiagent system for the study of urbanism. *Environment and Planning B*, 24, 287-306.

Scheiner, J., & Holz-Rau, C. (2013). Changes in travel mode use after residential relocation: a contribution to mobility biographies. *Transportation*, 40(2), 431-458.

Schirmer, P. M., van Eggermond, M. A., & Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2), 3-21.

Schneider, R. J. (2013). Theory of routine mode choice decisions: An operational framework to increase sustainable transportation. Transport Policy, 25, 128-137.

Schwarz, N., Kahlenberg, D., Haase, D., & Seppelt, R. (2012). ABMland-a tool for agent-based model development on urban land use change. *Journal of Artificial Societies and Social Simulation*, 15(2), 8.

Ševčíková, H., Raftery, A. E., & Waddell, P. A. (2007). Assessing uncertainty in urban simulations using Bayesian melding. *Transportation Research Part B: Methodological*, 41(6), 652-669.

Ševčíková, H., Raftery, A. E., & Waddell, P. A. (2011). Uncertain benefits: Application of bayesian melding to the alaskan way viaduct in seattle. *Transportation Research Part A: Policy and Practice*, 45(6), 540-553.

Shafiei, E., Stefansson, H., Asgeirsson, E. I., Davidsdottir, B., & Raberto, M. (2013). Integrated Agent-based and System Dynamics Modelling for Simulation of Sustainable Mobility. *Transport Reviews*, 33(1), 44-70.

Silva, C. (2013). Structural accessibility for mobility management. *Progress in Planning*, 81, 1-49.

Silva, E., & Wu, N. (2012). Surveying models in urban land studies. *Journal of Planning Literature*, 27(2), 139-152.

Silva, E. A. (2011). Cellular Automata and Agent Base Models for Urban Studies: From Pixels to Cells to Hexa-dpi's. *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, 323-334.

Silva, J. D. E., Martinez, L. M., & Goulias, K. G. (2012). Using a multi equation model to unravel the influence of land use patterns on travel behavior of workers in Lisbon. *Transportation Letters-the International Journal of Transportation Research*, 4(4),

Simmonds, D. (2001). The objectives and design of a new land-use modelling package:

DELTA Regional science in business, London: Springer.
Simmonds, D., & Still, B. (1999). DELTA/START: Adding land use analysis to integrated transport models. In World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research 4. Antwerp, Belgium, July 12–17, 1998
Simmonds, D., & Still, B. (1999). DELTA/START: Adding land use analysis to integrated transport models. *World Transport Research, Vols* 1 to 4, C1-C14.
Simmonds, D. C. (1999). The design of the DELTA land-use modelling package. *Environment and Planning B*, 26, 665-684.
Simon, D., McGregor, D., & Nsiah-Gyabaah, K. (2004). The changing urban-rural interface of African cities: definitional issues and an application to Kumasi, Ghana. *Environment and Urbanization*, 16(2), 235-248.
Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1), 99-118.
Simon, H. A. (1957). *Models of Man: Social and Rational*. Oxford, England: Wiley.
Simon, H. A. (2000). Bounded rationality in social science: today and tomorrow. *Mind & Society*, 1(1), 25-39.
Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional Science and Urban Economics*, 40(5), 292-298.
Sinai, I. (2001) Moving or Improving: Housing Adjustment Choice in Kumasi, Ghana, *Housing Studies*, 16(1), 97- 114
Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of economic literature*, 332-382.
Sun, Y. L., Waygood, E. O. D., Fukui, K., & Kitamura, R. (2009). Built Environment or Household Life-Cycle Stages-Which Explains Sustainable Travel More? Case of Kyoto-Osaka-Kobe, Japan, Built Area. *Transportation Research Record* (2135), 123-
Hagerstrand T. (1970) What about people in regional science? *Papers of the Regional Science Association* 24: 1–12.
Tannier, C., Hirtzel, J., Stephenson, R., Couillet, A., Vuidel, G., & Youssoufi, S. (2016). Conception and use of an individual-based model of residential choice in a planning decision process. Feedback from an experimental trial in the city of Besançon, France. *Progress in Planning*, 108, 1-38.
Thiele, J. C., Kurth, W., & Grimm, V. (2012). Agent-Based Modelling: Tools for Linking NetLogo and R. *JASSS-the Journal of Artificial Societies and Social Simulation*, 15(3).
Thiele, J. C., Kurth, W., & Grimm, V. (2014). Facilitating parameter estimation and

sensitivity analysis of agent-based models: A cookbook using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, 17(3), 11.

Tilahun, N., & Levinson, D. (2011). Work and home location: Possible role of social networks. *Transportation Research Part a-Policy and Practice*, 45(4), 323-331.

Tilahun, N., & Levinson, D. (2013). Selfishness and altruism in the distribution of travel time and income. *Transportation*, 40(5), 1043-1061.

Tilahun, N., & Levinson, D. (2013). An Agent-Based Model of Origin Destination Estimation (ABODE). *Journal of Transport and Land Use*, *6*(1), 73-88.

Tipple, A. G., Korboe, D. T., & Garrod, G. D. (1997). A comparison of original owners and inheritors in housing supply and extension in Kumasi, Ghana. *Environment and Planning B: Planning and Design*, 24(6), 889-902.

Tobias, R., & Hofmann, C. (2004). Evaluation of free Java-libraries for social-scientific agent based simulation. *Journal of Artificial Societies and Social Simulation*, 7(1).

Torrens, P. M. (2007). A geographic automata model of residential mobility. Environment and Planning B: Planning and Design, 34(2), 200-222.

Torrens, P. M. (2012). Moving Agent Pedestrians Through Space and Time. *Annals of the Association of American Geographers*, 102(1), 35-66.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76(1), 31.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281.

Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis*, 9(2), 163-173.

Veldhuisen, J., Timmermans, H., & Kapoen, L. (2000). RAMBLAS: a regional planning model based on the microsimulation of daily activity travel patterns. *Environment and Planning A*, 32(3),

Von Bertalanffy, L. (1950). An outline of general system theory. British Journal for the Philosophy of Science.

Von Bertalanffy, L. (1993). General system theory. George Braziller. Traduction française: Théorie générale des systèmes, Dunod.

Waddell, P. (1993). Exogenous Workplace Choice in Residential Location Models: Is the Assumption Valid? *Geographical Analysis*, 25(1), 65-82.

Waddell, P. (2000). A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. *Environment and Planning B*, 27(2), 247-264.

Waddell, P. (2000). Towards a behavioral integration of land use and transportation modelling. In *International Conference on Travel Behaviour Research, 9th, 2000, Gold Coast, Queensland, Australia, Vol 1*.

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association*, 68(3), 297-314.

Waddell, P., Wang, L., Charlton, B., & Olsen, A. (2010). Microsimulating parcel-level land use and activity-based travel: Development of a prototype application in San Francisco. *The Journal of Transport and Land Use*, 3, 65–87.

Waddell, P. (2011). Integrated land use and transportation planning and modelling: addressing challenges in research and practice. *Transport Reviews*, 31(2), 209-229.

Waddell, P., Bhat, C., Eluru, N., Wang, L., & Pendyala, R. M. (2007). Modeling interdependence in household residence and workplace choices. *Transportation Research Record: Journal of the Transportation Research Board*, 2003(1), 84-92.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., & Ulfarsson, G. (2003). Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics*, 3(1), 43-67.

Waddell, P. A. (1997). Household choice and urban structure: a reassessment of the behavioral foundations of urban models of housing, labor and transportation markets: Ashgate Aldershot.

Wagner, P., & Wegener, M. (2007). Urban land use, transport and environment models: experiences with an integrated microscopic approach. *disP-The Planning Review*, 43(170), 45-56.

Wan, X., Zhang, Y., Jin, P. J., Ran, B., Wang, W., & Chen, J. (2011). Same-Day Mode Choice Modeling with Household Vehicle Usage Simulation in Developing Countries. *Transportation Research Record* (2239), 23-33.

Wang, D. G., Chai, Y. W., & Li, F. (2011). Built environment diversities and activity-travel behaviour variations in Beijing, China. *Journal of Transport Geography*, 19(6), 1173-1186.

Wang, D. G., & Lin, T. (2013). Built environments, social environments, and activity-travel behavior: a case study of Hong Kong. *Journal of Transport Geography*, 31, 286-295.

Wang, L. M., Waddell, P., & Outwater, M. L. (2011). Incremental Integration of Land Use and Activity-Based Travel Modeling Workplace Choices and Travel Demand. *Transportation Research Record* (2255), 1-10.

Wang, M., Qian, Y. S., & Guang, X. P. (2012). Improved calculation method of shortest path with cellular automata model. *Kybernetes*, 41(3-4), 508-517.

Webster, D. (2002). On the edge: Shaping the future of peri-urban East Asia: Asia/Pacific Research Center Stanford.

Wegener, M. (1982). Modeling urban decline: a multilevel economic-demographic model for the Dortmund region. *International Regional Science Review*, 7(2), 217-241.

Wegener, M. (2004). Overview of land-use transport models. *Handbook of transport geography and spatial systems*, 5, 127-146.

Wegener, M. (2011). From macro to micro—how much micro is too much? *Transport Reviews*, 31(2), 161-177.

Wegener, M., & Fürst, F. (2004). Land-use transport interaction: state of the art. Url: http://econwpa.repec.org/eps/urb/papers/0409/0409005.pdf. [Accessed 28th September 2014]

Weng, F. L., & Duan, J. (2009). Analysis of Land Use-Transport Interaction Model Research. Proceedings of the International Conference on Urban and Regional Planning in Transitional China and the 3rd International Association for China Planning Conference, 217-222.

Wilensky, U. (1999). NetLogo. https://ccl.northwestern.edu/netlogo/

Wilensky, U., & Rand, W. (2007). Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation*, 10(4), 2.

Wilensky, U., & Rand, W. (2015). An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo: MIT Press.

Wilson, A. G. (1970). Entropy in urban and regional modelling: Pion Ltd.

Wingo, L. (1961). Transportation and urban land. Washington, DC: Resources for the Future. Inc., l96l, 255.

Wu, N., & Silva, E. A. (2010). Artificial intelligence solutions for urban land dynamics: a review. *Journal of Planning Literature*, 24(3), 246-265.

Xie, Y., Batty, M., & Zhao, K. (2007). Simulating emergent urban form using agent-based modeling: Desakota in the Suzhou-Wuxian region in China. *Annals of the A ssociation of American Geographers*, 97(3), 477-495.

Yang, J., & Ferreira Jr, J. (2008). Choices versus choice sets: A commuting spectrum method for representing job - Housing possibilities. *Environment and Planning B: Planning and Design*, 35(2), 364-378.

Yang, J. W., French, S., Holt, J., & Zhang, X. Y. (2012). Measuring the Structure of U.S.

Metropolitan Areas, 1970-2000 Spatial Statistical Metrics and an Application to Commuting Behavior. *Journal of the American Planning Association*, 78(2), 197-209.

Yang, L. Y., Zheng, G., & Zhu, X. N. (2013). Cross-nested logit model for the joint choice of residential location, travel mode, and departure time. *Habitat International*, 38, 157-166.

Yasmin, F., Morency, C., & Roorda, M. J. (2016). Macro-, meso-, and micro-level validation of an activity-based travel demand model. Transportmetrica A: *Transport Science*, 1-28.

Yin, R. K. (2013). Case study research: Design and methods, London: Sage publications.

Zegras, C. (2010). The Built Environment and Motor Vehicle Ownership and Use: Evidence from Santiago de Chile. *Urban Studies*, 47(8), 1793-1817.

Zhang, H. H. I., Zeng, Y. N. A., Bian, L., & Yu, X. J. (2010). Modelling urban expansion using a multi agent-based model in the city of Changsha. *Journal of Geographical Sciences*, 20(4), 540-556.

Zhang, Y., Zhang, X., Zheng, M., & Long, Y. (2017). An Integrated Model of Transportation and Land Use for Development and Application in Beijing. *International Review for Spatial Planning and Sustainable Development*, 5(1), 71-85.

Zhao, L. Y., & Peng, Z. R. (2010). Integrated Bilevel Model to Explore Interaction Between Land Use Allocation and Transportation. *Transportation Research Record* (2176), 14-25.

Zhao, P. J. (2010). Sustainable urban expansion and transportation in a growing megacity: Consequences of urban sprawl for mobility on the urban fringe of Beijing. *Habitat International*, 34(2), 236-243.

Zhao, P. J. (2011). Car use, commuting and urban form in a rapidly growing city: evidence from Beijing. *Transportation Planning and Technology*, 34(6), 509-527.

Zhao, P. J. (2013). The Impact of the Built Environment on Individual Workers' Commuting Behavior in Beijing. *International Journal of Sustainable Transportation*, 7(5), 389-415.

Zhou, B., Kockelman, K. M., & Lemp, J. D. (2009). Applications of Integrated Transport and Gravity-Based Land Use Models for Policy Analysis. *Transportation Research Record* (2133), 123-132.

Zhou, Y. X., & Wang, X. (2012). Overview of Computer Simulation Modeling Platforms Based on Multi-Agent. *Advanced Manufacturing Technology*, Pts 1-4, 472-475, 3111-3116.

Zhu, W., & Timmermans, H. (2010). Cognitive process model of individual choice behaviour incorporating principles of bounded rationality and heterogeneous decision heuristics. *Environment and Planning B-Planning & Desig*n, 37(1), 59-74.

Zhuge, C., Shao, C., Gao, J., Dong, C., & Zhang, H. (2016). Agent-based joint model of residential location choice and real estate price for land use and transport model. *Computers, Environment and Urban Systems*, 57, 93-105.

Zolfaghari, A., Sivakumar, A., & Polak, J. W. (2012). Choice set pruning in residential location choice modelling: a comparison of sampling and choice set generation approaches in greater London. *Transportation Planning and Technology*, 35(1), 87-106.